

**APLICACIÓN DE UN MODELO DE SOBREVIDA PARA EL ANÁLISIS DE LA  
DESERCIÓN EN LOS PROGRAMAS DE INGENIERÍA ELECTROMECÁNICA Y  
DISEÑO INDUSTRIAL DE LA UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA  
DE COLOMBIA**

**YINA LISSETH ALVAREZ PALACIOS**

**UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA  
FACULTAD SECCIONAL DUITAMA  
LICENCIATURA EN MATEMÁTICAS Y ESTADÍSTICA  
DUITAMA  
2016**

**APLICACIÓN DE UN MODELO DE SOBREVIDA PARA EL ANÁLISIS DE LA  
DESERCIÓN EN LOS PROGRAMAS DE INGENIERÍA ELECTROMECÁNICA Y  
DISEÑO INDUSTRIAL DE LA UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA  
DE COLOMBIA**

**YINA LISSETH ALVAREZ PALACIOS**

**TRABAJO DE GRADO EN MODALIDAD MONOGRAFÍA**  
**Para optar por el título de Licenciada en Matemáticas y Estadística**

**DIRECTORA DEL PROYECTO**  
**CARMEN HELENA CEPEDA ARAQUE**  
**MSc. Estadística**

**UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA**  
**FACULTAD SECCIONAL DUITAMA**  
**LICENCIATURA EN MATEMÁTICAS Y ESTADÍSTICA**  
**DUITAMA**  
**2016**

Nota de aceptación

---

---

---

---

---

---

Presidente del Jurado

MSc. CLARA EMILCE ROJAS MORALES

---

Jurado

MSc. SANDRA PATRICIA CÁRDENAS

---

Jurado

MSc. REINALDO ALARCÓN GUARÍN

Duitama, 29 de Junio 2016

*Dedico este trabajo en primer lugar a  
DIOS, por ser siempre aquel que reafirmó  
en cada momento la tranquilidad,  
persistencia y serenidad necesaria en  
esta etapa de la vida que esta próxima a  
culminar.*

## **AGRADECIMIENTOS**

Agradezco inmensamente a Dios por guiarme y darme la sabiduría necesaria para afrontar cada reto que se presentó en esta linda carrera universitaria y así dar lo mejor de mí para salir adelante.

A mi madre, por ser un ejemplo de mujer trabajadora, persistente y luchadora ante cada adversidad también por su constante ánimo, apoyo y compañía que siempre está dispuesta a brindarme sin importar las circunstancias.

A mi novio Camilo Araque, un compañero incondicional, quien ha compartido esta última etapa de mis estudios, porque en su compañía las cosas malas se convierten en buenas además por estar siempre a mi lado como un sabio consejero.

A la directora de este proyecto de grado, la profesora MSc. Carmen Helena Cepeda Araque, por toda la paciencia, dedicación y el gran esfuerzo realizado durante el desarrollo del trabajo, deseo expresar mi gratitud hacia usted deseándole éxito en su vida y trayectoria profesional.

A mis compañeros Víctor Camargo y Juan Carlos Aguilar, por brindarme su compañía en este proceso, me queda la satisfacción de haber compartido con personas tan valiosas como ustedes, les doy las gracias por su apoyo y afecto.

En general le agradezco a todas y cada una de las personas que me han ayudado con su apoyo, colaboración y amistad durante el transcurso de mi carrera y especialmente en este proyecto.

***YINA LISSETH ALVAREZ PALACIOS***

**TÍTULO:** APLICACIÓN DE UN MODELO DE SOBREVIDA PARA EL ANÁLISIS DE LA DESERCIÓN EN LOS PROGRAMAS DE INGENIERÍA ELECTROMECÁNICA Y DISEÑO INDUSTRIAL DE LA UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA

**AUTORA:** ALVAREZ PALACIOS, Yina Lisbeth.

**DIRECTORA:** CEPEDA ARAQUE, Carmen Helena.

**PUBLICACIÓN:** Duitama. Universidad Pedagógica y Tecnológica de Colombia, 2016.

**UNIDAD PATROCINANTE:** Universidad Pedagógica y Tecnológica de Colombia, Facultad Seccional Duitama, Escuela de Matemáticas y Estadística.

**PALABRAS CLAVES:** Deserción estudiantil universitaria, modelos de sobrevivida, regresión de Cox.

**OBJETIVO:**

Establecer, a través de un modelo de sobrevivida, las variables determinantes en la disminución del riesgo de deserción y del aumento en los niveles de graduación de los estudiantes del programa de Ingeniería Electromecánica como el de Diseño Industrial de la Facultad Seccional Duitama.

**DESCRIPCIÓN:**

Este proyecto que se presenta busca responder a la siguiente pregunta de investigación: ¿Cuáles son las variables que aparecen como fuertes determinantes en la disminución del riesgo de deserción y del aumento en los niveles de graduación de cada uno de los programas de Ingeniería Electromecánica y Diseño Industrial de la Facultad Seccional Duitama de la Universidad Pedagógica y Tecnológica de Colombia?

El cual se desarrolla a través de unos objetivos específicos, los cuales se mencionan como sigue:

Establecer el marco conceptual de la deserción estudiantil en la Facultad Seccional Duitama y, dependiendo de la disponibilidad de información, determinar las variables académicas y no académicas que se utilizarán para su medición en el tiempo.

Identificar el momento de inicio y la longitud del proceso de monitoreo de los estudiantes de los Programas Académicos, a partir de la información reportada en bases de datos de la Universidad.

Evaluar las variables académicas y no académicas relacionadas con la deserción mediante la construcción de un modelo de sobrevivencia que permita cuantificar la incidencia de cada una de ellas y determinar las más influyentes.

Usar la función de riesgo del modelo de sobrevivencia para cada uno de los estudiantes activos en I semestre de 2015 y calcular la probabilidad de abandono voluntario de sus estudios en el segundo semestre de 2015.

Diseñar un cuestionario de Excel para recoger información de variables no estáticas que permita a largo plazo construir un modelo de sobrevivencia.

#### **FUENTES:**

Para el desarrollo del proyecto se consultaron 4 libros de estadística, específicamente de modelos de sobrevivencia, además se consultaron 2 artículos, 1 diapositiva sobre modelos de sobrevivencia, en los cuales se describían estudios referentes a la deserción universitaria y los factores que influyen sobre este fenómeno. Para el caso de la obtención del marco muestral se acudió a los sistemas de SPADIES (Sistema para la Prevención de la Deserción en las Instituciones de Educación Superior) y se complementó la medición con la información brindada por la Oficina de Registro y Control Académico de Tunja, particularmente la base de datos SIRA (Sistema de Información y Registro Académico). El diseño muestral correspondió a un muestreo aleatorio simple sin reposición, en este muestreo se realiza la selección de un subconjunto de elementos de la población en forma aleatoria sin reposición.

#### **CONTENIDO:**

El documento consta de seis partes, empezando por una presentación del proyecto, continuando con un marco referencial en el cual se muestra antecedentes investigativos, conceptualizaciones de deserción y la teoría de los modelos de

sobrevida aplicados al fenómeno de la deserción, en tercer lugar se tiene el diseño metodológico establecido en el proyecto. Una cuarta y quinta parte se refiere a los modelos de sobrevida en cada programa académico de estudio, donde se presenta un análisis descriptivo de variables y en seguida la construcción del modelo ajustado. Finalmente se dan unas conclusiones y recomendaciones.

### **METODOLOGÍA:**

Las fases metodológicas que permitirán la descripción y explicación de la naturaleza longitudinal del proceso de deserción, a través de un modelo de sobrevida, en los programas de Ingeniería Electromecánica y Diseño Industrial de la Universidad Pedagógica y Tecnológica de Colombia –Duitama, serán:

1. Conformación del marco teórico el cual partirá de la revisión documental sobre la deserción, antecedentes investigativos y modelos de sobrevida.
2. Definición del diseño metodológico, el cual, y a partir de la disponibilidad de información en bases de datos de la Universidad, como por ejemplo el Sistema de Información y Registro Académico (SIRA) implica la selección de los variables estáticas académicas (tales como aumento de los cupos, política de cancelaciones, desajustes del calendario académico por paros, por ejemplo) y estáticas no académicas (tales como ingreso de la familia, puntaje icfes, sexo, nivel educativo de la madre, estado de empleo cuando presentó el icfes, propiedad de la vivienda, edad presentación del icfes, número de personas en la familia, estrato, nivel SISBEN). También esta fase implica determinar el momento de inicio y la longitud del proceso de monitoreo de los estudiantes.
3. Depuración de la información y ajuste para aplicación de la técnica estadística.
4. Construcción del modelo estadístico.
5. Conclusiones y elaboración del informe final del proyecto de grado.

### **CONCLUSIONES:**

Como conclusión principal del programa de Diseño Industrial se tiene que el 60% de los estudiantes que ingresan desertan, cifra que según el MEN (2014) está por encima del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2% e incluso la del área básica de



conocimiento del Programa – Ingeniería, arquitectura, urbanismo y Afines - que se encuentra en 50,7%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 60, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables. Para el caso de DI se identificó que el 26.25% de los estudiantes se gradúan.

La mayoría deserta por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Deserta en su mayoría los hombres, características sobresalientes de los estudiantes que ingresaron a Diseño Industrial son la edad promedio de ingreso (17 años), el 65% de los estudiantes contaban con vivienda propia, el 84% no trabajaban al momento de presentar el examen de estado, el puntaje promedio del examen de estado fue de 67.6, el 35,6% de los estudiantes tenían madres con nivel educativo de básica primaria y el 44,5% de los estudiantes tenían dos hermanos al momento de presentar el examen de Estado.

Para el programa de DI, la posibilidad de que un estudiante se vaya en primer semestre es del 19%, donde el periodo crítico con mayor intensidad es de manera temprana (del primero al tercer semestre). Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas.

Al determinar qué variables están asociadas con el estado del estudiante (desertor o no desertor) se determinó que sólo ocurre con la variable “trabajaba a la hora de presentar el examen de estado también llamado prueba SABER 11” y que el semestre en que deserta y el puntaje en las pruebas de estado están correlacionados positivamente, es decir, a medida que aumenta el puntaje también aumenta el tiempo hasta que un estudiante se va del programa.

El factor que aumenta el riesgo de deserta en DI es trabajar al momento de presentar su examen de estado, lo cual se puede atribuir a la falta de tiempo con las actividades académicas, es decir que en la medida que un estudiante no trabaje se reduce el riesgo de deserción, se recomienda a la Universidad y al programa crear políticas de patrocinio y apoyo a aquellos estudiantes con esta característica. Se disminuye el riesgo de deserta al presentar mayor edad al ingresar al programa, lo cual puede evidenciarse debido al tema vocacional en los estudiantes que ingresan muy jóvenes a DI.

Por el contrario, para la graduación el factor que disminuye esta propensión es mayor número de asignaturas perdidas, luego un estudiante presenta menor probabilidad de graduarse. Además a partir del semestre número catorce es más probable que un estudiante se gradúe en el 14, hecho que es preocupante ya que gasta cuatro semestres más de lo previsto para obtener el título profesional y posiblemente entre más tiempo transcurra será difícil culminar con el proyecto educativo. Por lo anterior se sugirieron actividades para aumentar los niveles de graduación inmediatamente se culmine con la respectiva carga académica, por ejemplo la asignación de temas de grado en semestres inferiores, vinculación inmediata con algún programa de posgrado según la normatividad de la Universidad, o en la pasantía involúcralos con proyectos productivos.

En el programa de Ingeniería Electromecánica se identificó que el 43.57% de los estudiantes que ingresan desertan, cifra que según el MEN (2014) está por debajo del nivel universitario para Colombia que fue del 44.9% para el año 2013, respecto a la de Boyacá para ese mismo año es ligeramente superior pues está en 43.2% al comparar con el área básica de conocimiento del Programa– Ingeniería, arquitectura, urbanismo y Afines - también es inferior ya que se encuentra en 50,7%. De 100 estudiantes que ingresan al Programa lo abandonan 44, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

Se puede afirmar que la mayoría de estudiantes que ingresan a este programa son hombres (90.5%), el puntaje promedio obtenido en el examen de estado fue 89.39 en una escala de 0 a 100, el 75.97% contaban vivienda propia. Se resalta que dentro de la muestra ningún estudiante trabajaba. Además el 53.63% de hogares de los estudiantes tenían un ingreso entre 1 y 2 Salarios Mínimos Mensuales Legales Vigentes (SMMLV) al momento de presentar el examen de estado, otra variable relacionada con este examen es el número de hermanos, el 44.69% de los estudiantes tenían dos hermanos en ese momento, respecto al nivel educativo de las madres el mayor porcentaje es 39.1% correspondiente a básica secundaria. El literal C del Artículo 80 del reglamento estudiantil es la mayor causa en la deserción académica con un 44.12% del total de los desertores por motivos académicos, por otro lado la deserción por causas no académicas fue obtenida en su gran mayoría por el retiro definitivo con un 84.1% que hace referencia a los estudiantes que pasaron carta de retiro definitivo del programa.

El 40.22% de los estudiantes que ingresaron entre el 2004 y 2009 se graduaron y el 43.57% desistieron de estudiar Ingeniería Electromecánica, predominando la deserción temprana (entre el segundo y quinto semestre), el literal C del Artículo 80 del reglamento estudiantil fue el mayor motivo de deserción académica y la deserción por causas no académicas tiene como mayoría a los estudiantes que se retiran voluntariamente del Programa.

Para el programa, la posibilidad de que un estudiante se vaya en primer semestre es del 12%, donde el periodo crítico, es decir con mayor intensidad es de manera temprana (del primero al quinto semestre). Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas.

Para ING, el modelo de sobrevivencia para la deserción permitió identificar que a partir de las variables consideradas en este estudio no se puede detectar aquellas que inciden en tal fenómeno, por lo cual se recomienda analizar otras variables para un próximo estudio. Para la graduación los factores que disminuyen son la tenencia de vivienda, es decir los estudiantes que carecen de vivienda tiene una mayor probabilidad de graduarse, estudiantes con madres de nivel educativo de básica primaria al momento de presentar el examen de estado respecto a un estudiante con madre de otro nivel educativo y un caso contrario es el número de hermanos pues a mayor número de hermanos menor probabilidad de graduarse.

A partir de los resultados obtenidos se puede apreciar que los factores estudiados que afectan la deserción y la graduación (excepto el número de materias repetidas) son exógenos, difícilmente pueden ser cambiados por la universidad. Es decir los riesgos se asocian más con cuestiones de tipo estructural de nuestra sociedad. Así las cosas, es importante que la UPTC encamine sus esfuerzos a mecanismos alternativos de estudio que le permita a estudiantes con ciertas características “desfavorables” reducir los riesgos de deserción y aumentar sus posibilidades de graduación.

Es importante analizar la graduación en ambos programas de estudio pues según lo encontrado la probabilidad de terminar décimo semestre y no graduarse es muy alta, es después del semestre doce cuando inicia el aumento de la probabilidad de graduarse.

Se recomienda construir a futuro nuevos modelos de sobrevida a partir de la recolección de información a través del instrumento, el cual está disponible en el “Anexo D”, y que tiene como propósito recolectar la información sobre aspectos importantes que no fueron abordados en este proyecto.

En este estudio se usó para construir la función de sobrevida técnicas no paramétricas, sería conveniente en los trabajos posteriores asumir una distribución de probabilidad para las variables tiempo hasta la deserción y graduación.

## CONTENIDO

	Pág.
1. PRESENTACIÓN DEL PROYECTO .....	19
2. MARCO REFERENCIAL.....	25
2.1. ANTECEDENTES INVESTIGATIVOS .....	25
2.2. DESERCIÓN .....	25
2.3. TEORÍA DE LOS MODELOS DE SOBREVIDA .....	32
2.3.1. Características de los datos de sobrevivida .....	33
2.3.2. Especificación del modelo de sobrevivida .....	33
2.3.3. Estimación del modelo de sobrevivida .....	42
2.4. MODELO DE RIESGOS PROPORCIONALES .....	46
2.4.1. Estimación de parámetros del modelo de COX .....	47
2.4.2. Selección del modelo.....	48
2.4.3. Evaluación del modelo.....	51
3. DISEÑO METODOLÓGICO.....	52
3.1. DISEÑO MUESTRAL.....	55
3.2. RECOLECCIÓN DE LOS DATOS .....	55
3.3. VARIABLES OBJETO DE ESTUDIO .....	57
3.4. TÉCNICAS DE ANÁLISIS DE DATOS.....	61
4. MODELO DE SOBREVIDA PARA DISEÑO INDUSTRIAL.....	63
4.1. DESCRIPCIÓN DE LOS DATOS.....	63
4.2. MODELO PARA DESERCIÓN.....	66
4.2.1. Función de sobrevivida .....	71
4.2.2. Modelo de COX – Factores que influyen en la deserción .....	71
4.2.3. Evaluación del modelo de COX .....	78
4.2.4. Pronóstico a partir del modelo .....	82
4.3. MODELO PARA GRADUACIÓN.....	83
4.3.1. Función de sobrevivida .....	85
4.3.2. Modelo de COX – Factores que influyen en la graduación .....	86
4.2.3. Evaluación del modelo de COX .....	88

5. MODELO SOBREVIDA PARA INGENIERIA ELECTROMECAÁNICA .....	95
5.1. DESCRIPCIÓN DE LOS DATOS.....	95
5.2. MODELO PARA DESERCIÓN.....	101
5.2.1. Función de Sobrevida.....	101
5.2.2. Modelo de Cox – Factores que influyen en la deserción.....	103
5.3. MODELO PARA GRADUACIÓN.....	108
5.3.1. Función de Sobrevida.....	108
5.3.2. Modelo de Cox – Factores que influyen en la graduación.....	108
5.3.3. Evaluación del modelo de Cox.....	110
6. CONCLUSIONES Y RECOMENDACIONES.....	114
BIBLIOGRAFÍA .....	123
ANEXOS.....	127

## LISTA DE GRÁFICAS

	Pág.
Gráfica 1. Diagrama de Lexis para DI .....	71
Gráfica 2. Función de sobrevida estimada para deserción en DI .....	71
Gráfica 3. Función de riesgo acumulado para deserción de DI .....	73
Gráfica 4. Función de sobrevida por género en DI .....	74
Gráfica 5. Función de sobrevida por trabajaba en DI .....	75
Gráfica 6. Otras curvas de sobrevida de DI.....	77
Gráfica 7. Función de sobrevida para el modelo de deserción en DI.....	80
Gráfica 8. Función de riesgo acumulado del modelo de Cox de deserción en DI .	81
Gráfica 9. Residuos de Cox-Snell para deserción en DI.....	82
Gráfica 10. Residuos escalonados de Schoenfeld de deserción en DI.....	83
Gráfica 11. Función de sobrevida estimada para graduación en DI .....	86
Gráfica 12. Función de riesgo acumulado para graduación de DI .....	87
Gráfica 13. Función de sobrevida para el modelo de graduación en DI .....	90
Gráfica 14. Función riesgo acumulado del modelo de Cox de graduación en DI..	91
Gráfica 15. Residuos de Cox-Snell para graduación en DI .....	92
Gráfica 16. Residuos escalonados de Schoenfeld de graduación en DI .....	92
Gráfica 17. Diagrama de Lexis para ING.....	102
Gráfica 18. Función de sobrevida para ING .....	103
Gráfica 19. Función de riesgo acumulado para ING.....	104
Gráfica 20. Otras funciones de sobrevida en ING .....	105
Gráfica 21. Función de sobrevida para nivel educativo madre en ING .....	106
Gráfica 22. Función de sobrevida para graduación de ING .....	109
Gráfica 23. Función de riesgo acumulado para graduación en ING .....	110
Gráfica 24. Función estimada para el modelo de graduación en ING.....	113
Gráfica 25. Función de riesgo acumulado para graduación en ING .....	114
Gráfica 26. Residuos de Cox-Snell para graduación en ING .....	115
Gráfica 27. Residuos de Cox-Snell para graduación en ING .....	116

## LISTA DE TABLAS

	Pág.
Tabla 1. Comparación de curvas de sobrevida .....	46
Tabla 2. Parámetros para determinar el tamaño de la muestra .....	56
Tabla 3. Listado de variables objeto de estudio.....	60
Tabla 4. Resumen descriptivo de las variables de estudio de DI.....	63
Tabla 5. Resumen bivariado de las variables de estudio de DI .....	66
Tabla 6. Test de correlaciones de Pearson entre tiempo en desertar y variables cuantitativas en DI .....	69
Tabla 7. Características de los datos para el modelo de sobrevida de DI .....	70
Tabla 8. Resumen porcentaje de censurados DI.....	70
Tabla 9. Estimaciones función de sobrevida para deserción en DI.....	72
Tabla 10. Resumen estimador Kaplan Meier por género de DI .....	74
Tabla 11. Resumen estimador Kaplan Meier por trabajaba en DI .....	76
Tabla 12. Modelo de Cox para la deserción en DI.....	78
Tabla 13. Resumen modelo óptimo para la deserción en DI .....	79
Tabla 14. Estimación de los parámetros del modelo de Cox de deserción en DI .	80
Tabla 15. Estimaciones función de sobrevida del modelo de deserción en DI ....	81
Tabla 16. Riesgos proporcionales de deserción en DI .....	82
Tabla 17. Riesgo relativo de deserción en DI .....	85
Tabla 18. Estimaciones función de sobrevida para graduación en DI .....	87
Tabla 19. Modelo de Cox para graduación en DI .....	88
Tabla 20. Resumen modelo óptimo para graduación en DI .....	89
Tabla 21. Estimación del modelo de Cox de graduación en DI .....	89
Tabla 22. Estimaciones función de sobrevida del Modelo de graduación en DI ...	90
Tabla 23. Riesgos proporcionales de graduación en DI .....	91
Tabla 24. Resumen descriptivo de las variables de estudio de ING .....	95
Tabla 25. Resumen bivariado de las variables de estudio del programa de ING	99
Tabla 26. Test de correlaciones de Pearson entre tiempo en desertar y variables cuantitativas de ING .....	100
Tabla 27. Características de los datos para el modelo de sobrevida .....	101
Tabla 28. Resumen porcentaje de censurados para ING .....	102
Tabla 29. Estimaciones función de sobrevida para ING .....	104
Tabla 30. Estimaciones para Ingresos familiares en ING .....	107



Tabla 32. Estimaciones por Kaplan Meier para graduación en ING .....	109
Tabla 33. Modelo de Cox para graduación en ING .....	110
Tabla 34. Resumen modelo óptimo para graduación en ING .....	111
Tabla 35. Estimación del modelo de Cox de graduación en ING.....	113
Tabla 36. Estimaciones función de sobrevida para el modelo de graduación en ING .....	114
Tabla 37. Riesgos proporcionales en graduación de ING .....	115

## TABLA DE ANEXOS

	Pág.
ANEXO A. CÓDIGOS ESTUDIANTES MUESTRA DE ING .....	127
ANEXO B. CÓDIGOS ESTUDIANTES MUESTRA DE DI .....	128
ANEXO C. SENTENCIAS DE R.....	129
ANEXO D. FORMATO DE REGISTRO DE DATOS.....	144

## 1. PRESENTACIÓN DEL PROYECTO

La deserción a nivel universitario es un problema que ha sido abordado desde diferentes puntos de vista, la mayoría de trabajos son de tipo descriptivo, pero con el fin de crear estrategias de prevención ante este fenómeno y mejorar la inversión de los recursos de las universidades se necesita identificar el riesgo de abandono, pues algunos estudiantes no cumplen con sus proyectos de educación.

Con este fin se evidencia la necesidad de usar una herramienta para el análisis del fenómeno de la deserción diferente a las comúnmente usadas, por esta razón en este proyecto se propone el uso del análisis de sobrevivencia con lo cual se permitirá calcular el periodo más probable en que un estudiante de los programas del estudio decida desertar, también se identifican las causas que aumentan ese riesgo de deserción.

Cifras como que el 45% de los estudiantes que ingresan a educación superior, no completan sus estudios hasta graduarse (Educación de Calidad, el camino para la prosperidad, 2010-2014), justifican el desarrollo de proyectos como el de “Ampliación de la cobertura en la Educación Superior”, del Ministerio de Educación Nacional, que busca incrementar la cobertura apoyándose en acciones, entre otras, como la de promover en las Instituciones de Educación Superior (IES) acciones orientadas a fomentar la permanencia de los estudiantes en el sistema, aumentando su capacidad de monitorear e implementar estrategias para disminuir la deserción, en este sentido el Ministerio de Educación Nacional –MEN- menciona “la educación superior en los últimos años se ha caracterizado por aumentos de cobertura e ingreso de estudiantes nuevos, el número de alumnos que logra culminar sus estudios es bajo, dejando entrever que una gran parte de estos los abandona, primordialmente en los primeros semestres” (Ministerio de Educación Nacional, 2007).

La deserción estudiantil genera tres tipos de problemas, el primero de ellos es el estudiante, ya que por los factores de superación y posicionamiento económico genera frustración, el segundo es la universidad ya que pierde individuos y dinero porque invierte sus recursos en personas que abandonan sus proyectos, y por último la sociedad en donde se genera un incremento del subempleo.

Para brindar soluciones al respecto el Ministerio de Educación Nacional ha liderado junto con las instituciones de educación superior, el diseño y la operación de una metodología de seguimiento de la deserción estudiantil en educación superior que

se concreta en el Sistema para la Prevención y Análisis de la Deserción en las Instituciones de Educación Superior (SPADIES). Esta herramienta permite además de hacer un análisis del riesgo de deserción, identificar y clasificar a los estudiantes en riesgo de desertar, dado un grupo determinado de variables. Estadísticamente hablando, lo que permite el SPADIES es estimar el riesgo de desertar en un punto determinado del tiempo (Ministerio de Educación Nacional, 2007, pág. 40), sin embargo, hasta el año 2009, aproximadamente, el análisis de la deserción se había abordado desde un marco estático ignorando la evolución del fenómeno a lo largo del tiempo. Es decir, se describía por qué un estudiante decide abandonar sus estudios pero no se explicaba el proceso de abandono. Específicamente, el tiempo que ha durado el estudiante hasta el momento de desertar.

Lo dinámico de los modelos de sobrevida, permite determinar el riesgo de ocurrencia de un evento, en este caso el evento de desertar, y analizar cuándo es más probable que éste ocurra teniendo en cuenta la influencia de sus principales factores o predictores.

A partir de esta perspectiva, el Ministerio de Educación Nacional, con el fin de contribuir a que las instituciones de educación superior dispongan de información para el diseño de políticas antideserción, elaboró un estudio denominado “Deserción estudiantil en la educación superior colombiana” con el propósito de contribuir a entender el problema de la deserción a través de su análisis en un modelo de sobrevida y como guía para el diagnóstico, diseño de acciones y evaluación de las mismas en las instituciones de educación superior (Ministerio de Educación Nacional, 2007, pág. 17).

Por supuesto el fenómeno de la deserción no es ajeno a la Universidad Pedagógica y Tecnológica de Colombia ubicada en el departamento de Boyacá. Institución que cuenta con cuatro sedes, 50 programas presenciales y 21 a distancia, de los cuales siete programas presenciales pertenecen a la seccional que se encuentra en la ciudad de Duitama, facultad interdisciplinar que cuenta con tres programas en administración: Turística y Hotelera, Empresas Agropecuarias e Industrial; dos programas de licenciatura uno en Tecnología y otro en Matemáticas y Estadística; uno de Ingeniería Electromecánica y uno en Diseño Industrial.

La Universidad, derivando del Plan de Desarrollo Institucional, Lineamiento Formación y Docencia, ha venido ejecutando el programa “Permanencia y Deserción”. Programa que busca caracterizar cualitativa y cuantitativamente la deserción para definir las causas de este fenómeno y establecer acciones para

minimizarla, tales como el Plan Padrino, el sistema de becas, monitores, y los programas de bienestar. Con estas acciones se trata de disminuir la deserción.

Además de lo señalado en el tema de la deserción universitaria, este trabajo permite, en primer lugar, al estudiante de Licenciatura en Matemáticas y Estadística poner en práctica su perfil ocupacional de asesor estadístico en proyectos, ya que consiste en la aplicación de una técnica de modelamiento estadístico para analizar un tema relevante para un educador, como lo es el de deserción. La técnica que se utilizará es un modelo de sobrevida, y resulta particularmente novedoso ya que a la fecha la deserción en los programas de la Facultad no ha sido abordada a través de ésta técnica estadística de dependencia.

En segundo lugar, el objeto de estudio que plantea esta monografía es el fenómeno de la deserción de los estudiantes constituido en un problema sobre el cual debe actuar la Universidad, y que está directamente relacionado con la eficiencia de sus recursos y el cumplimiento de sus fines misionales. De hecho, el tema que se abordará resulta de gran interés para la comunidad académica, ya que mediante la interpretación de los parámetros estimados de un modelo de sobrevida, se proporciona a las autoridades académicas de la Facultad Seccional Duitama información sustentada y necesaria para formular posibles políticas o reformas educativas que logren aumentar la permanencia de los estudiantes dentro de la Institución.

Es claro que todos los estudiantes que abandonan una institución crean lugares vacantes que pudieron ser ocupados por alumnos que persistieran en los estudios. La pérdida de estudiantes causa problemas financieros a las instituciones al producir inestabilidad en la fuente de sus ingresos, pero también un alto costo social que puede asociarse a la pérdida de productividad laboral derivada de la menor acumulación individual de capital humano.

En tercer lugar, tal como lo plantea el estudio denominado “Deserción estudiantil en la educación superior colombiana”, la deserción es un problema que afecta la relación del Estado con las instituciones de educación superior públicas, en el sentido del incumplimiento de las políticas y las metas sociales establecidas. Aspecto que resalta la relevancia de describir amplia y profundamente el comportamiento de la deserción en los programas de Ingeniería Electromecánica y Diseño Industria de la Facultad.

En cuarto lugar, es evidente que el estudio de fenómenos como la deserción permite aportar elementos sustanciales en los procesos de autoevaluación, autorregulación y mejoramiento continuo de los Programas Académicos que se estudiarán.

Ya que el proyecto busca obtener una descripción y explicación de la naturaleza longitudinal del proceso de deserción, a través de un modelo de sobrevida, pues dicha técnica de análisis permite seguir la variable dependiente hasta que ocurra el evento de interés, es decir la deserción. En consecuencia, este proyecto busca responder a la siguiente pregunta de investigación:

¿Cuáles son las variables que aparecen como fuertes determinantes en la disminución del riesgo de deserción y del aumento en los niveles de graduación en los programas de Ingeniería Electromecánica y Diseño Industrial de la Universidad Pedagógica y Tecnológica de Colombia?

El proceso que se utilizará para la solución de la pregunta establecida está ligado al *objetivo general* que se planteó: “establecer, a través de un modelo de sobrevida, las variables determinantes en la disminución del riesgo de deserción y del aumento en los niveles de graduación de los estudiantes del programa de Ingeniería Electromecánica como el de Diseño Industrial de la Facultad Seccional Duitama”. Este objetivo se cumple por medio de los siguientes objetivos específicos:

Establecer el marco conceptual de la deserción estudiantil en la Facultad Seccional Duitama y, dependiendo de la disponibilidad de información, determinar las variables académicas y no académicas que se utilizarán para su medición en el tiempo.

Identificar el momento de inicio y la longitud del proceso de monitoreo de los estudiantes de los Programas Académicos, a partir de la información reportada en bases de datos de la Universidad.

Evaluar las variables académicas y no académicas relacionadas con la deserción mediante la construcción de un modelo de sobrevida que permita cuantificar la incidencia de cada una de ellas y determinar las más influyentes.

Usar la función de riesgo del modelo de sobrevida para cada uno de los estudiantes activos en I semestre de 2015 y calcular la probabilidad que de abandono voluntario de sus estudios en el segundo semestre de 2015.

Diseñar un cuestionario de Excel para recoger información de variables no estáticas que permita a largo plazo construir un modelo de sobrevivencia que las incluya tales como relaciones intra-familiares y los problemas entre el estudiante y los miembros de su núcleo familiar; desintegración de los hogares por muerte de progenitores, separación o divorcio; migración al exterior, a otras regiones; el estado de salud, tanto física como mental del estudiante, enfermedades crónicas, incapacitantes, minusvalía; factores socioeconómicos; estudiantes con hijos, embarazos no planeados, matrimonio; situaciones del contexto político social: persecución política, desplazamiento forzado, amenazas, exilio entre otros; situaciones casuales como fallecimiento, calamidad doméstica.

## **2. MARCO REFERENCIAL**

Para la construcción del marco referencial, se han tenido en cuenta varios documentos tanto del tema de deserción universitaria como sobre los modelos de sobrevivencia y el modelo de riesgos proporcionales; a continuación se nombran y describen algunos estudios, que conforman la base de esta monografía.

### **2.1. ANTECEDENTES INVESTIGATIVOS**

Según el Ministerio de Educación Nacional (MEN), se destacan tres trabajos, el primero, fue realizado entre la Universidad Nacional de Colombia y el Instituto Colombiano para el Fomento de la Educación Superior (ICFES), cuyo resultado fue una contribución a la comprensión teórica y conceptual de la deserción; el segundo, realizado por la Universidad de Antioquia, en el que se implementaron técnicas estadísticas adecuadas para estudiar la deserción como un problema dinámico, apoyado por encuestas y, el tercero, desarrollado por el Ministerio de Educación Nacional con apoyo de la Universidad de los Andes, en el cual se hizo un análisis del riesgo de deserción y además crearon el Sistema de Prevención de la Deserción en Educación Superior o SPADIES, que permite a cada institución identificar y clasificar a los estudiantes en riesgo de deserción, dado un grupo determinado de variables.

Diversos estudios se han realizado con el objetivo de estudiar el fenómeno de la deserción a través de modelos de sobrevivencia. Algunos de estos estudios se describen a continuación ya que sirven como antecedentes teóricos para la monografía que se presenta.

Osorio, Bolancé, & Castillo, (2011) en su artículo deserción universitaria, resaltan dos corrientes literarias: el modelo de integración del estudiante (Student integration Model) Tinto (1975) y Spady (1970), en el cual se plantea que entre mayor sea el grado de integración al ambiente académico y social, es menor la probabilidad de abandonar los estudios; el modelo de desgaste del estudiante (Bean, 1980), a través de él se destaca el efecto que tienen las variables externas a la institución, como por ejemplo los factores individuales.

Según las autoras Tinto (1989) afirma que el estudio de la deserción en la educación superior es extremadamente complejo, ya que implica no sólo una variedad de perspectivas sino también una gama de diferentes tipos de abandono.



Con el modelo de integración del estudiante, las autoras citan a Robinson (1990) que presenta la deserción como una relación directa con la interacción entre estudiante-profesor y también estudiante-compañeros, obteniendo de esta forma que “basados en este tipo de modelo, Radcliffe, Huesman y Kellog (2006) plantean que las variables que mejor predicen la probabilidad y el momento de desertar son el rendimiento académico en el primer semestre y la preparación académica” (p. 34). En el modelo de desgaste del estudiante, mencionan que “Willet y Singer (1991) plantean la probabilidad de desertar mayor para estudiantes con menor rendimiento académico y que tienen padres de menor nivel educativo y menores ingresos” (p. 34).

Las autoras también mencionan en el mismo artículo, estudios que combinan los dos modelos anteriores, un ejemplo es el Instituto Colombiano para la Evaluación de la Educación (ICFES) y la Universidad Nacional de Colombia (2002), lograron determinar que el género, la edad, las condiciones académicas y económicas, son bastante significativas, tanto para explicar la deserción como el rezago y la graduación.

En el documento “Modelos de desercion universitaria” Osorio, Bolancé & Castillo (2009) mencionan que la Universidad de los Andes, con ayuda del Centro de Estudios sobre Desarrollo Económico (CEDE) (2006), utilizan información de las instituciones de educación superior de Colombia, encontrando que los factores más sobresalientes que aumentan la elección sobre decidir desertar corresponden a, ser hombre, la educación de los padres, la ocupación laboral, el bajo puntaje en el examen de ingreso a la educación superior (SABER 11), estudiar programas como ingeniería, arquitectura y ciencias de la educación, estudiar en una universidad privada y la falta de un apoyo financiero.

Por otro lado, Castaño Velez, y otros, (2009) empiezan su análisis de la deserción, destacando la teoría del suicidio de Durkheim (1897) que se compara con el suicidio en la sociedad, donde los factores socioeconómicos, la guerra, la religión entre otros formaban parte de las explicaciones para el fenómeno del suicidio, el llamado suicidio anómico (ocurre cuando una falla lleva a la desorientación individual), definición dada por Palacio (2010, págs. 6-7); estos autores también mencionan a Spady del modelo de integración social.

Por su parte, (Giovangnoli, 2002) utiliza un modelo de duración para identificar los factores que explican la deserción estudiantil y la graduación; dentro de sus principales conclusiones la autora plantea que el tipo de colegio, la educación de

los padres, el género y la situación laboral del estudiante son los factores de mayor relevancia para determinar la probabilidad de que un estudiante abandone o no la universidad.

Una forma de comprender la deserción es comparándola con un abandono que puede ser explicado por diferentes variables: socioeconómicas, individuales, institucionales, académicas, entre otras. Vale la pena comentar que del texto se resalta la descripción de las relaciones que se ven afectadas por la deserción.

Es importante señalar que las instituciones educativas intentan identificar el motivo de abandono (académico o no académico) que influencia a los estudiantes a retirarse de estudiar, con ello determinar las acciones a seguir, en este punto, radican los cambios evolutivos que han tenido los estudios de este problema, todos con el fin de explicar la deserción. Por lo anterior el Ministerio de Educación Nacional (MEN) inició la revisión de la literatura existente en el año 2003 como se menciona en el documento Deserción estudiantil en la educación superior colombiana, el objetivo era construir el estado del arte sobre la deserción estudiantil; a partir de los resultados obtenidos en el estudio se crearon estrategias para la retención de estudiantes teniendo como ayuda principal el hecho de centrarse en las variables determinantes de la disminución del riesgo de deserción y el aumento de los niveles de graduación.

A continuación se presentan algunos estudios sobre deserción en Colombia:

En el año 2003 la Universidad de Antioquia (Colombia) hizo unas encuestas a estudiantes desertores de la cohorte de 1996 pertenecientes a las facultades de ingeniería y ciencias económicas, se aplicó el modelo de regresión de riesgo proporcional en tiempo discreto de Prentice-Gloeckler y Meyer, añadiendo al último una variable aleatoria gamma, se destacan como resultados que estudiantes que no recibieron orientación vocacional tienen un mayor riesgo de deserción como aquellos que tienen personas a cargo pues este factor económico es también influyente; las principales características de los estudiantes desertores encontradas en el modelo son, estar soltero y vivir de manera independiente.

Castaño Velez, y otros, (2009) en el documento Deserción Estudiantil en la Educación Superior de Colombia mencionan el estudio llamado causas e indicadores de la deserción en el programa de economía cohortes desde 1997 hasta 2005 de la universidad del Atlántico tenía como objetivo aplicar los modelos de duración y macroeconómico (2006), con los estudiantes del programa.

Barrera Rebellón (2008) efectuó un análisis de supervivencia aplicado al problema de la deserción estudiantil en la Universidad Tecnológica de Pereira, se encontró, que dentro de los programas de ingenierías los estudiantes que provienen de colegios privados tienen una ligera superioridad en cuanto al nivel de supervivencia de los estudiantes que provienen de colegios públicos, además se menciona que en los programas de tecnologías, ingenierías y licenciaturas las mujeres presentan mayor nivel de supervivencia respecto de los hombres.

La Universidad Distrital Francisco José de Caldas, en el año 2011 realizó un estudio de retención de estudiantes por medio de su oficina de planeación, se pretendía identificar cuál era la situación de retención en la universidad y también valorar las consecuencias socioeconómicas del problema mediante su incidencia en el estudiante y la universidad, los resultados obtenidos son que por ejemplo el tiempo de graduación de un estudiante es de cuatro semestres adicionales para obtener su título, en el primer semestre de 2011 había un 25% de población de estudiantes de pregrado se encuentran cursando semestres adicionales, al hacer la valoración económica de la demora por parte de los estudiantes en promedio de cuatro semestres se obtuvo que gastan 56 y 79 salarios mínimos legales vigentes (SMLV) para el ciclo tecnológico y profesional respectivamente, por parte de la universidad a nivel de presupuesto se multiplicó el costo anual por estudiante y por el número de estudiantes, obteniendo una cifra cercana a los 26.000 millones.

Viloria y Peralya realizaron un análisis de los factores determinantes de la deserción en la universidad Popular del Cesar, estudio realizado con el fin de averiguar cuando se tiene más riesgo de desertar en la universidad e identificar las características socioeconómicas y los factores relacionados en la presencia de ese evento, se determinó que los hombres, los estudiantes con hijos, estudiantes matriculados con menos de tres materias y los estudiantes que trabajan tienen mayor probabilidad de desertar.

A partir de los anteriores antedecentes teóricos, se concluye que las teorías que delimitan esta monografía son modelo de integración del estudiante (Tinto, 1975 y Spady, 1970) y la teoría del desgaste del estudiante (Bean, 1980), pues con las variables académicas, institucionales, personales y socioeconómicas que estos autores mencionan, se realiza una mayor ampliación en la descripción del fenómeno de la deserción.

## 2.2. DESERCIÓN

Esta sección contiene varios aspectos, en primer lugar se presentará la definición de deserción universitaria establecida por el Ministerio de Educación Nacional (MEN), luego se describen los tipos de deserción y algunas características del tema, después se mencionan el reglamento estudiantil de la UPTC, y para concluir finalmente, se establece la definición de deserción tomada para este proyecto.

En el documento Deserción Estudiantil en la Educación Superior de Colombia del Ministerio de Educación Nacional (MEN) mencionan entre otros a Tinto y Giovagnoli, a partir de ellos, los autores Castaño Velez, y otros, (2009) dan una definición para la deserción: entendiéndola como la situación que enfrenta un estudiante cuando aspira y no logra concluir su proyecto educativo, por lo cual se considera como desertor al individuo que siendo estudiante de una institución de educación superior no presenta ninguna actividad académica durante dos semestres académicos consecutivos<sup>1</sup>.

En algunas investigaciones esa inactividad es llamada *primera deserción* pues no se conoce si el individuo volverá a estudiar esos estudios o iniciará otro.

Como afirman Castaño Velez, y otros, (2009) las estadísticas del Ministerio de Educación Nacional son alarmantes como por ejemplo, de cada cien estudiantes que ingresan a una institución de educación superior cerca de la mitad no logran culminar su ciclo académico ni obtener la graduación.

Es importante conocer las características de los estudiantes que están en riesgo de desertar, con un análisis superficial y sencillo se infiere que esos estudiantes en peligro presentan características similares a las de los desertores, aunque por diferentes motivos se mantienen en las aulas.

Entre las principales características de deserción se destacan: el bajo aprovechamiento de oportunidades educativas, ausencia de clases, problemas de indisciplina, falta de motivación, nivel socio-económico, inapetencia por el conocimiento, problemas relacionados con la edad. (Correa & Páramo, 1999)

---

<sup>1</sup> Definición de deserción tomada por el Ministerio de Educación Nacional (MEN).

Según el documento “Deserción estudiantil en la educación superior colombiana”, del Ministerio de Educación Nacional (MEN), se plantean dos tipos de clasificación para la deserción universitaria, respecto al tiempo y espacio en que se presenta.

La deserción con respecto al tiempo:

1. Deserción precoz: Individuo que siendo admitido por la institución de educación superior no se matricula, en el presente trabajo también se entenderá el abandono de sus estudios al finalizar el primer semestre.
2. Deserción temprana: Individuo que abandona sus estudios antes de cursar el 50 % de los semestres del programa.
3. Deserción tardía: Individuo que abandona los estudios después de cursar el 50% de los semestres del programa.

Existen distintas causas generadoras de la deserción, una de ellas es debido a las diferentes características de los estudiantes, por ello el Ministerio de Educación Nacional (MEN) plantea que existen dos periodos con mayor riesgo, el primero es cuando el estudiante tiene su primer contacto con la institución y el segundo hace referencia a los primeros semestres, es decir la adaptación al sistema social y académico. Aunque en los demás semestres también se presenta la deserción esta tiende a disminuir ya que el hecho de abandonar los estudios representa la pérdida de tiempo y dinero para el estudiante.

La deserción con respecto al espacio, se divide en:

1. Deserción institucional: Caso en el cual el estudiante abandona la institución.<sup>2</sup>
2. Deserción interna o del programa académico: Se refiere al alumno que decide cambiarse a otro programa que ofrece la misma institución de educación superior.

Dentro de las investigaciones mencionadas por el Ministerio de Educación Nacional (MEN) se destaca que la mayoría de literatura de este tema está focalizada sobre la parte institucional y se señalan que los cuatro conjuntos de variables (institucionales, socioeconómicas, académicas y personales) son estudiados por

---

<sup>2</sup> Existen varias posibilidades:

a) Estudiante que decide voluntariamente trasladarse de institución y se vincula a otra institución; y  
b) Estudiante que se retira voluntariamente de la institución y adicionalmente se desvincula del sistema de educación.

separado y como un grupo de variables que pueden ser determinantes al momento de decidir desertar.

Cabe mencionar que como señala el Ministerio de Educación Nacional (MEN), el trabajo de DesJardins, Ahlburg y MacCall (1999), es una técnica de análisis de historia de eventos, se abre una brecha en comparación con los anteriores estudios mencionados donde en general se describe el por qué un estudiante toma la decisión de desertar pero no explica el proceso del abandono. En ese mismo documento se explica la nueva técnica de análisis así: “Con el enfoque de historia de eventos, por el contrario, se obtiene una descripción y explicación de la naturaleza longitudinal del proceso de deserción, ya que el método de análisis permite seguir la variable dependiente hasta que ocurra el evento de interés; este tipo de análisis también se conoce como análisis de supervivencia, de duración o de modelación de riesgo” (p. 28)

Por otro lado, se resalta que la Universidad Nacional de Colombia efectuó un estudio en el año 2006 y encontró que dentro de los factores no socioeconómicos que aumentan la deserción, se encuentran las relaciones intra-familiares y problemas con miembros del núcleo familiar, estudiantes con hijos (embarazos no planeados), situaciones del contexto político social (persecuciones políticas, desplazamiento, amenazas), la madurez del estudiante frente a su vida, estado de salud tanto física como mental.

La Universidad Pedagógica y Tecnológica de Colombia cuenta con su reglamento estudiantil mediante el acuerdo 130 de 1998, el cual contempla en el Artículo 80 las distintas formas de perder la calidad de estudiante como lo son por bajo rendimiento académico y demás situaciones que ameritan la no renovación de la matrícula, las cuales se encuentran registradas así:

- “a) Quien obtenga durante cuatro (4) semestres, un Promedio Aritmético Acumulado inferior a tres cero (3.0).
- b) Quien teniendo un Promedio Aritmético Acumulado inferior a tres cero (3.0), obtenga un promedio aritmético semestral inferior a dos cero (2.0).
- c) Quien pierda una asignatura que curse en calidad de repitente siendo su Promedio Aritmético Acumulado inferior a tres cero (3.0). En el caso en que el Promedio Aritmético Acumulado sea igual o superior a tres cero (3.0), la podrá cursar por tercera y última vez.
- d) Quien pierda en un mismo período académico dos asignaturas que se cursan en calidad de repitente.

e) Quien pierda una asignatura que cursa por tercera vez.”

En el mismo artículo se realiza una aclaración, se exceptúan los estudiantes que hayan aprobado y cursado el 80% del total de créditos del respectivo estudio, en ese caso se otorga una oportunidad por el siguiente semestre. Si persiste, se continua con el proceso de no renovación de matrícula.

Así mismo en el artículo 42 del Acuerdo 130 de 1998, se encuentra señalado que los estudiantes que no realicen renovación de matrícula durante tres semestres académicos consecutivos pierden su cupo.

En esta monografía se determinó que para clasificar la deserción de un estudiante se tendrá en cuenta lo contemplado en el Reglamento Estudiantil de la universidad (Acuerdo 130 de 1998), de esta forma se presenta deserción académica y no académica, entendiendo por deserción académica, lo contemplado en el Artículo 80 del Reglamento Estudiantil (bajo rendimiento académico) como se indicó en esta sección y la deserción no académica consta de la categoría “Retiro Definitivo” refiriéndose a los estudiantes que les fue aprobada la solicitud de retiro definitivo del programa y la categoría de “No renovó matrícula” en el caso indicado en el Artículo 42 del Reglamento Estudiantil mencionado en el párrafo anterior.

Por lo tanto y según las clasificaciones de la deserción, para este proyecto sólo se analizará con respecto al tiempo, sin tener en cuenta el espacio en que se presente ya que se cuenta sólo con la trazabilidad del estudiante en los programas de Diseño Industrial e Ingeniería Electromecánica, respectivamente.

### **2.3. TEORÍA DE LOS MODELOS DE SOBREVIDA**

En esta sección se hace una descripción del análisis con datos estadísticos de sobrevida, definiciones, características y las distribuciones que son usadas en los distintos casos de los modelos de duración.

El nombre sobrevivencia es debido al hecho que los primeros usos fueron eventos donde la variable de interés era la muerte; un análisis de sobrevivencia, es una técnica estadística mediante la cual se estudia una variable que representa el tiempo hasta que ocurre un evento y su correlación con otras variables.

La importancia de la sobrevivencia se basa en la flexibilidad que maneja pues en el caso de realizar un estudio estadístico se tendrían como recursos, las técnicas de

análisis de varianza o los modelos de regresión, pero como menciona Castaño Velez, y otros, (2009) “los modelos de duración (también conocidos como análisis de sobrevivencia o supervivencia) han venido adquiriendo gran popularidad para el estudio de la deserción estudiantil debido a que permiten realizar un análisis dinámico del fenómeno”.

Los modelos de sobrevida se usan para determinar la probabilidad de ocurrencia de un evento, en este caso la deserción, con ello se conoce cuándo es más probable que se presente el evento y qué factores influyen.

### **2.3.1. Características de los datos de sobrevida**

En los análisis de sobrevida, menciona Godoy (2009), el objetivo se encuentra en un grupo o varios grupos de individuos, centrandose en un evento que se estudia también llamado falla, que ocurre en un tiempo llamado tiempo de falla.

Este mismo autor define el tiempo de supervivencia como el tiempo desde la entrada al estudio o estado inicial hasta el estado final o el tiempo transcurrido hasta la ocurrencia del evento de interés. Con el fin de determinar con mayor exactitud el tiempo de falla, es necesario establecer el tiempo de origen, una escala de medición y el significado claro de falla.

Algunos obstáculos que se presentan en el análisis de la sobrevivencia son la censura y el truncamiento, que poseen características diferentes, más adelante se explican los tipos de casos en cada uno.

Según Díaz (2015), la estructura con las características de los datos de sobrevida son la unidad sobre la cual se registra el evento, el evento de interés, el tiempo origen o inicial y el tiempo final del estudio, la escala de medida del tiempo hasta el evento, el tipo de censura y el truncamiento del tiempo.

#### **Tiempo de falla**

En el documento “Introducción al análisis de supervivencia con R”, Godoy (2009), señala la importancia de determinar el evento que se identifica como la falla desde el inicio del estudio, generalmente en los análisis de sobrevida la escala de medición usada es el tiempo real (días, meses, años), aunque también se usa el kilometraje o longitud hasta el primer defecto en textiles, por ejemplo.



En los estudios médicos, el tiempo de falla se podría tomar como el tiempo hasta la muerte del paciente por causas del estudio; en la industria, la falla es el tiempo hasta primer momento en el cual el desempeño, cae por debajo del nivel aceptable que se estableció. En el caso del proyecto, el tiempo de falla, es el semestre hasta que un estudiante deserta de alguno de los programas en estudio.

### **Censura**

La siguiente definición de censuramiento es la dada por Castaño Velez, y otros, (2009), en el documento del Ministerio de Educación Nacional (MEN): “considérese una población homogénea de individuos (estudiantes),  $i = 1, \dots, n$ , los cuales pueden experimentar el evento de interés (deserción) asumiendo que para cada individuo éste no es repetible, es decir, una vez que el evento ocurre no es posible que suceda nuevamente.”

En los análisis de sobrevivencia se llaman casos censurados cuando hay individuos a los cuales no les ha ocurrido el evento, por ejemplo, algunos estudiantes no presentan el evento de deserción porque cumplen con la terminación de los estudios.

Como se considerará la censura en algunos casos y dependiendo del caso en particular se trabaja con una función de verosimilitud diferente, por ello en los siguientes párrafos se identifican las características que pueden presentar los datos, a continuación se explican los tipos de censura (derecha, izquierda e intervalo). Godoy (2009) explica los tipos de censura de la siguiente forma:

### **Censura por la derecha**

En primera medida se considera la **Censura tipo I**, el evento es observado sólo si ocurre antes de un tiempo establecido, independientemente del tamaño de la muestra. La notación para este caso es:

$X$  = *Tiempo de vida para un individuo específico bajo estudio*

$C_r$  = *Tiempo fijo de censura*

$X$ 's = *Variables aleatorias independientes e idénticamente distribuidas*

$f(x)$  = *Función de densidad*

Así, el tiempo de vida exacto de un individuo es conocido si y solo si  $X \leq C_r$ .

Si  $X > C_r$ , el individuo es un sobreviviente y su tiempo de vida es censurado  $C_r$ .

El segundo caso es la **Censura tipo II**, en la cual hay dependencia del tamaño de la muestra (denotado por  $n$ ) y las fallas que se observen. Aquí todos los individuos son puestos en estudio al mismo tiempo y se da el término de este cuando  $r$  es un número entero positivo determinado previamente por el investigador, tal que  $r < n$ .

Otro tipo de Censura es la **Tipo III** o también llamada Censura aleatoria, la cual surge cuando los sujetos salen del estudio sin presentar la falla por razones no controladas por el investigador. Por ejemplo, en un estudio donde el evento de interés es la muerte por alguna razón ajena a la de interés o si el investigador pierde acceso al sujeto y este sale del estudio.

Para este proyecto solo se abordará la censura a derecha tipo I.

### **Censura por la izquierda**

Un tiempo de vida  $X$  asociado con un individuo específico en el estudio, se considera censurado por la izquierda, si este es menor que un tiempo de censura  $C_1$ . Esto es, que el evento de interés le ha ocurrido al sujeto en estudio, antes que el sujeto haya sido observado por el tiempo  $C_1$ .

### **Censura por intervalo**

Este es un tipo de censura más general que ocurre cuando el tiempo de vida se sabe que ocurre solamente dentro de un intervalo. Este tipo de censura se presenta cuando se tiene un estudio longitudinal donde el seguimiento del estado de los sujetos se realiza periódicamente y por ello, la falla sólo se conoce entre dos periodos de revisión, generando un intervalo para cada sujeto en el estudio.

### **Truncamiento**

El truncamiento es definido como una condición que presentan ciertos sujetos en el estudio y el investigador no puede considerar su existencia. Cuando los datos presentan truncamiento, solamente individuos a los que les ocurre algún evento particular, antes del evento de interés o la censura, son considerados en el análisis por el investigador.

### **Truncamiento por la derecha**

Se refiere a los casos donde los individuos que presentaron el evento son los únicos incluidos en la muestra y los sujetos que no haya presentado el evento, no serán considerados. Por ejemplo una muestra que presenta truncamiento por la derecha, es un estudio de mortalidad basado en los registros de muerte.

### Truncamiento por la izquierda

Este ocurre cuando los sujetos entran al estudio a un tiempo particular (no necesariamente el origen del evento de interés), y son observados desde este “tiempo retrasado de entrada”, hasta que el evento ocurra o hasta que el evento es censurado. Si  $Y$  es el momento de ocurrencia del evento que trunca a los sujetos en estudio, entonces para muestras truncadas por la izquierda, solo los individuos tales que  $Y$  serán considerados. El tipo más común de truncamiento por la izquierda ocurre cuando los sujetos entran al estudio a un tiempo aleatorio y son observados por este “tiempo retrasado de entrada”, hasta que el evento ocurre o hasta que el sujeto es censurado por la derecha.

### 2.3.2. Especificación del modelo de sobrevida

Es el tiempo de supervivencia o tiempo de vida, representado por  $T$ , variable aleatoria continua no negativa, que denota los tiempos de duración de los individuos de la población, para el caso de este proyecto corresponde al tiempo de permanencia del estudiante hasta el momento de desertar.

De una forma más específica Díaz (2015), señala los casos de la distribución de probabilidad:

$$P(T \leq t) = \begin{cases} \sum_{t_j \leq t} p(t_j) = F(t), \text{ caso discreto} \\ \int_0^t f(u) du = F(t), \text{ caso continuo} \end{cases}$$

El mismo autor señala las siguientes propiedades de esta función:

- Del Teorema Fundamental del Cálculo  $f(t) = -\frac{dS(t)}{dt}$
- $S(0) = 1$
- $S(\infty) = \lim_{t \rightarrow \infty} S(t) = \lim_{t \rightarrow \infty} \int_t^{\infty} f(u) du = 0$
- Si  $a < b$ , entonces  $S(a) - S(b) = \int_a^b f(u) du \geq 0$
- La función de sobrevivencia  $S$  es monótona no creciente.

La distribución de la variable  $T$  se caracteriza por cuatro funciones básicas, llamadas, función de sobrevida, función de riesgo, función de riesgo acumulado y por último la función de vida media residual; las cuales están relacionadas de forma

matemática, es decir, teniendo una de ellas se pueden conocer los valores de las otras.

### 2.3.2.1. Función de sobrevivencia

Esta función básica es empleada para describir un evento a través del tiempo, la función es denotada por  $S(t)$ , además también es llamada tasa de sobrevivencia acumulativa, la función indica la probabilidad de que un sujeto sobreviva al evento de interés antes de un momento dado, la variable aleatoria  $T$  no negativa (de tiempo de falla) posee función de distribución  $F(t)$  y función de densidad de probabilidad  $f(t)$ , las que a continuación se describen:

Función de distribución acumulada:  $F(t) = P(T \leq t)$ , indica la probabilidad de que el evento de interés ocurra en un tiempo menor o igual que  $t$ .

La función de sobrevivencia:  $S(t) = P(T > t)$   
 $= P(\text{un individuo sobreviva mas allá de } t).$

De otra forma:  $S(t) = 1 - F(t)$   
 $= 1 - P(T \leq t)$   
 $= 1 - P(\text{un individuo falle antes del tiempo } t).$

En la figura 1, se observan los dos casos de la función sobrevivencia según la variable, continua y discreta. Es evidente que en el caso continuo la función es curva, pero es diferente con variables discretos, la cual tiene forma escalonada.

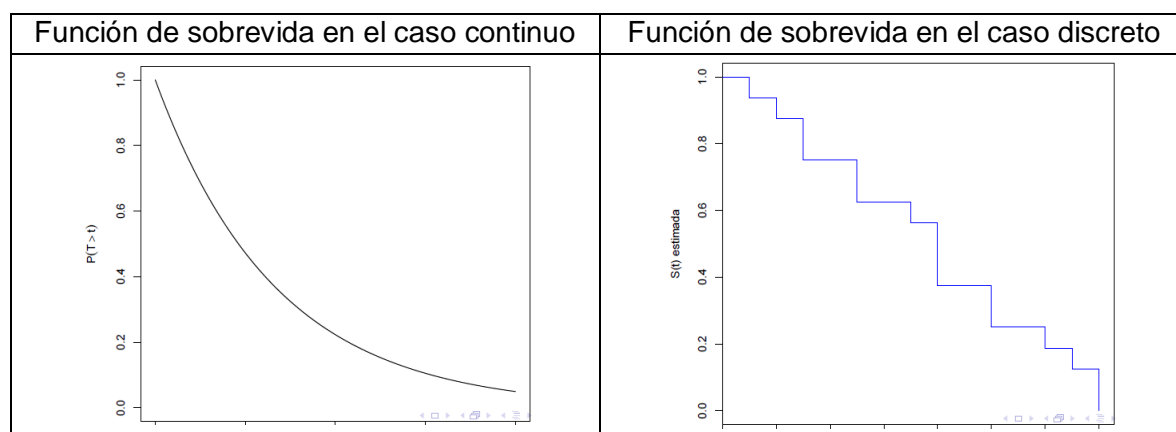


Figura 1. Gráficas de funciones de sobrevivencia

Fuente: Díaz (2015)

El tiempo de sobrevivencia  $T$ , por ser una variable aleatoria posee una función de densidad, esta función evalúa la probabilidad de que en una unidad el evento ocurra

después de un cierto tiempo  $t$  o equivalente, el evento no se ha presentado sobre la unidad antes del tiempo  $t$ .

En la figura 2, se tiene dos curvas de la curva de densidad de la función de sobrevivida, en la figura 2 (a) da un patrón de alta tasa de fallas al principio del estudio y una tasa decreciente de fallas cuando se incrementa el tiempo. En la figura 2 (b), el pico de frecuencia alta de fallas ocurre a aproximadamente 1.7 unidades de tiempo. La proporción de individuos que cae entre 1 y 2 unidades de tiempo es igual al área sombreada que aparece en la figura 2.

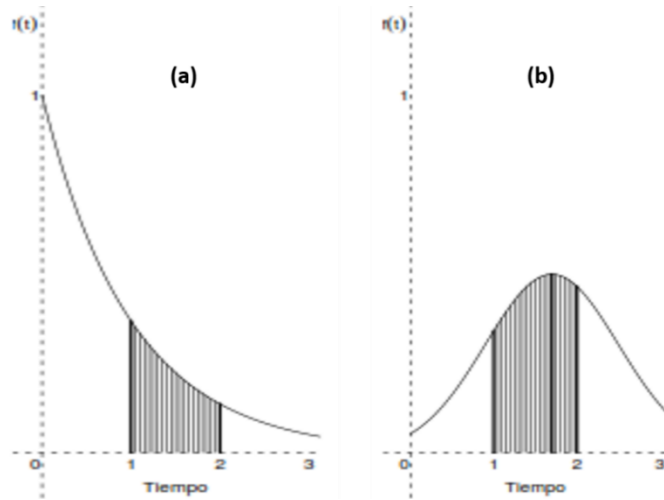


Figura 2. Curvas de densidad de la función de sobrevivida Fuente: Godoy (2009)

### 2.3.2.2. Función de riesgo

Otra función de interés es la llamada **función de riesgo**,  $h(t)$ , definida por Godoy (2009) como la probabilidad de falla en un intervalo muy pequeño de tiempo, teniendo en cuenta que el individuo de estudio ha sobrevivido hasta el inicio del intervalo, o como el límite de la probabilidad de que un individuo falle en un intervalo corto,  $(t, t+\Delta t)$ , dado que no ha ocurrido hasta el tiempo  $t$ . Esta función describe la tasa instantánea de falla a través del tiempo.

Díaz (2015) reduce la función de riesgo,  $h(t)$  para  $t>0$ , así:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T > t)}{\Delta t} \quad (1)$$

Desarrollando la probabilidad condicionada se obtiene:

$$h(t) = \frac{\lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t}}{P(T > t)} \quad (2)$$

Con el cálculo de probabilidades y utilizando  $P(T > t) = s(t)$ , se tiene así que:

$$h(t) = \frac{\lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}}{S(t)} = \frac{\frac{d(F(t))}{dt}}{S(t)} \quad (3)$$

De la definición de derivada se tiene que la expresión (3) se convierte en:

$$h(t) = \frac{f(t)}{S(t)} \quad (4)$$

Con la expresión (4), se tiene la relación entre la función de probabilidad, la función de sobrevivencia y la función de riesgo. Al realizar la integración de la expresión (4), se obtiene entonces la función de riesgo acumulada:

$$H(t) = \int_0^t h(w) dw = \int_0^t \frac{f(w)}{S(w)} dw \quad (5)$$

La cual se puede escribir así:  $H(t) = \int_0^t \frac{f(w)}{1 - F(w)} dw$

Realizando la siguiente sustitución  $u = 1 - F(w)$ , y  $du = -f(w)dw$

se tiene:  $H(t) = \int \frac{-du}{u}$ .

Colosimo & Ruiz Giolo (2006) determina las siguientes relaciones entre las cuatro funciones básicas antes mencionadas apartir de la ecuación (5):

$$H(t) = -\ln S(t) \quad (5.1)$$

$$S(t) = e^{-H(t)} \text{ con } H(t) = \int_0^t h(w) dw \quad (5.2)$$

$$h(t) = -d(\ln S(t))/dt \quad (5.3)$$

$$f(t) = -dS(t)/dt \quad (5.4)$$

Para ilustrar e identificar las funciones de riesgo se presentan algunos escenarios indicados por Godoy (2009), en la figura 3:

Pacientes con leucemia que no responden al tratamiento tienen una tasa de riesgo creciente  $h_1(t)$ .

La función de riesgo  $h_2(t)$  es decreciente pues indica el riesgo de soldados heridos por bala que fueron sometidos a cirugía. El peligro principal es la operación misma y éste decrece si la cirugía es exitosa.

Una función de riesgo constante como en  $h_3(t)$  es el riesgo de individuos saludables entre 18 y 40 años de edad cuyos riesgos principales de muerte son los accidentes.

La llamada curva de tina de baño  $h_4(t)$  que describe el proceso de vida humana, durante el periodo inicial el riesgo es alto (alta mortalidad infantil), luego el riesgo permanece casi constante hasta un cierto tiempo, después del cual vuelve a crecer debido a las fallas del organismo por la tercera edad.

Pacientes con tuberculosis tienen riesgos que se incrementan inicialmente, decrecen después de tratamiento, se muestra en la función de riesgo  $h_5(t)$ .

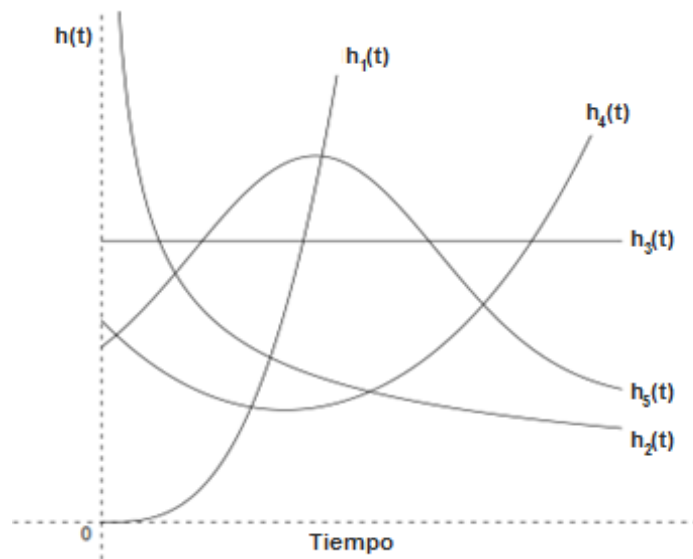


Figura 3. Representación de algunas gráficas de funciones de riesgo

Fuente: Godoy (2009)

Para el caso discreto, se tiene que siendo  $T$  una variable aleatoria discreta que toma valores  $t_j$  con  $j = 1, 2, \dots$ , la función de riesgo está definida para los valores  $t_j$  y proporciona la probabilidad condicional de falla al tiempo  $t = t_j$ , dado que el individuo estaba vivo antes de  $t_j$ , por lo tanto se tiene que:

$$h(t_j) = P(T = t_j | T \geq t_j) = \frac{P(T = t_j)}{P(T \geq t_j)} = \frac{f(t_j)}{S(t_j-)}$$

Donde  $t_j -$  corresponde a un instante antes de  $t_j$  y por lo tanto

$$P(T \geq t_j) = 1 - P(T < t_j) = S(t_j-) \neq S(t_j) \text{ caso discreto}$$

Luego de realizar algunos procedimientos matemáticos, Godoy (2009) determina a  $h(t)$  así:

$$h(t) = \prod_{t_j \leq t} \frac{S(t_j)}{S(t_{j-1})} \quad (6)$$

### 2.3.2.3. Función de Riesgo Acumulado

La función de riesgo acumulado se denota por  $H(t)$  y según Godoy (2009) en su documento Introducción al análisis de supervivencia con R, corresponde a:

$$H(t) = \int_0^t h(u) du \text{ caso continuo,}$$

$$H(t) = \sum_{t_j \leq t} h(t_j) \text{ caso discreto}$$

Aunque se presenta un inconveniente con la relación  $S(t) = \exp[-H(t)]$  Ya que en el caso discreto no es correcta, pero se usa como una aproximación, debido a que:

$$S(t) = \exp\{-H(t)\} = e^{h(t_1)} e^{h(t_2)} \dots e^{h(t_j)} \text{ con } t_j \leq t.$$

Que no corresponde con la relación que hay entre  $S(t)$  y  $h(t)$  en el caso discreto. .Por lo cual se define a la función de riesgo acumulado para el caso discreto como:

$$H(t) = -\sum_{t_j \leq t} \ln[1 - h(t_j)]. \quad (7)$$

Ecuación bien definida dado que  $0 < h(t_j) < 1$ , según

$$h(t_j) = 1 - \frac{S(t_j)}{S(t_{j-1})}$$



y para los valores  $t_j$  donde  $S(t_j)$  tiene sentido en el caso discreto, sucede lo siguiente:  $S(t_j) > S(t_{j+1})$

De tal modo que

$$S(t) = \exp\{-H(t)\} = \exp\left\{\sum_{t_j \leq t} \ln[1 - h(t_j)]\right\} = \prod_{t_j \leq t} (1 - h(t_j)). \quad (8)$$

Esta última expresión concuerda con la relación entre  $S(t)$  y  $h(t)$  en el caso discreto.

Por último Godoy (2009) en el mismo documento menciona “en ambos casos, tanto el discreto como el continuo, esta función como su nombre lo indica, acumula el riesgo al paso del tiempo. De tal manera que corresponde a una función no decreciente y de acuerdo a su forma de incrementarse, se podrá tener información del comportamiento del riesgo a lo largo del tiempo, lo cual es una ventaja en el análisis de supervivencia.” (pág. 29)

#### 2.3.2.4. Función de Vida Media Residual

Otra función en los análisis de sobrevivencia, es vida media residual al tiempo  $t_0$  denotada por  $mrl(t_0)$  (por el nombre en inglés *mean residual life*). Para los sujetos de edad  $t_0$ , la función se encarga de medir la esperanza del tiempo de vida restante, o el tiempo esperado antes de la ocurrencia del evento de interés. Díaz (2015) define esta función así:

$$mrl(t) = E(T - t | T > t). \quad (9)$$

En el caso continuo, por la definición de esperanza condicional se tiene que

$$\begin{aligned} E(T - t_0 | T > t_0) &= \int_{t_0}^{\infty} (t - t_0) dP(T \leq t | T > t_0) \\ &= \int_{t_0}^{\infty} (t - t_0) d\left(\frac{S(t_0) - S(t)}{S(t_0)}\right) \end{aligned}$$

Con lo anterior la función de *vida media residual* al tiempo  $t_0$  se define por:

$$mrl(t) = E(T - t_0 | T > t_0) = \frac{\int_{t_0}^{\infty} S(t) dt}{S(t_0)} \quad (10)$$

Por esto la vida media residual es el área bajo la curva de supervivencia a la derecha de  $t$  dividida entre  $S(t)$ .

De acuerdo con ello se tiene la vida media  $\mu = E(T) = E(T - t_0 | T > t_0) = mrl(0)$   
Lo anterior señala el área total de la curva de sobrevida.

Las distribuciones que se usan en el análisis de supervivencia son adecuadas para ciertas aplicaciones, según Colosimo & Ruiz Giolo cada una de estas distribuciones de probabilidad puede generar estimadores diferentes para la misma cantidad desconocida. Algunas de las principales distribuciones de probabilidad usadas en el análisis de supervivencia son la distribución Exponencial, Weibull, Log-normal, Gamma y Gamma generalizada.

De acuerdo a lo mencionado en el anterior párrafo la elección de la distribución de probabilidad resulta de vital importancia, pues se generan distintos estimadores, pues con una buena elección se evita usar un modelo inadecuado que no será acertado en sus conclusiones.

Como mencionan Colosimo & Ruiz Giolo existen otras distribuciones como lo son logística, log-logística, log-gama, rayleigh, normal inversa y gompertz; los periodos generales para los que se usa la función de tasa son:

Periodo de fallas prematuras o mortalidad infantil  
Periodo de vida útil  
Periodo de desgaste

### **2.3.3. Estimación del modelo de sobrevida**

En esta sección se mencionan las distintas estimaciones que se pueden realizar en un modelo de sobrevida dependiendo de las características, como menciona Godoy (2009) “algunos tiempos de falla pueden ser caracterizados por familias de distribuciones específicas que solo dependen de uno o varios parámetros desconocidos, los cuales proporcionan las características específicas del modelo en estudio”, en otros casos no es así, por ello los modelos de sobrevida pueden ser paramétricos (distribución de probabilidad) y no paramétrico, los cuales más

adelante se describen, así como la estimación por medio del estimador de Kaplan Meier, la comparación de funciones y por último la estimación del riesgo acumulado.

Los modelos de probabilidad según Colosimo & Ruiz Giolo, (2006) se presentaron para cantidades desconocidas (parámetro); la distribución gama generalizada tiene tres parámetros, las distribuciones Weibull, log-normal y gama tiene dos parámetros y la exponencial tan solo uno; todo esto se estima a partir de las observaciones de la muestra.

Para la estimación de los modelos en regresión lineal se usa frecuentemente el método de mínimos cuadrados, una razón para no usarlo en los estudios de tiempo de vida es la incapacidad de poder tener en cuenta las censuras, por ello es más apropiado el método de máxima verosimilitud, ya que es simple e incorpora las censuras, es relativamente más sencillo de entender y posee propiedades óptimas para muestras grandes.

Al realizar la estimación no paramétrica para modelos de sobrevida no es necesario suponer un modelo probabilístico para la variable tiempo hasta el evento. Cuando se tiene la presencia de datos censurados, el análisis descriptivo (media, desviación estándar, etc.) y técnicas gráficas (histograma, box-plot, etc.) no son adecuadas pues la división del tiempo en un cierto número de intervalos y las fallas en número de ocurrencias en cada intervalo no consideran la censura.

Entre los test que no son paramétricos se destaca el de Kaplan Meier por la comparación de dos o más funciones de supervivencia, el cual se describe en seguida.

#### **2.3.3.1. Estimación de la función de Sobrevida por Kaplan Meier**

El estimador de Kaplan–Meier, de la función de supervivencia es un estimador común entre los paquetes estadísticos, el cual no hace supuestos sobre la forma o distribución de las funciones de sobrevida, de riesgo y además, lo que si tiene en cuenta son las observaciones censuradas y no censuradas.

Según Díaz (2015) en el estimador de Kaplan-Meier se observan los tiempos al evento  $t_1, t_2, \dots, t_n$ . Algunos pueden ser censurados por la derecha y puede suceder que más de una unidad presente el mismo tiempo para el evento. Sea  $r$  el número de tiempos, donde  $r \leq n$ . Los  $r$  tiempos de falla ordenados en forma ascendente son  $t_{(1)}, t_{(2)}, \dots, t_{(r)}$ ; el  $j$ -ésimo tiempo hasta el evento es denotado por  $t_{(j)}$  teniendo

$j = 1, 2, \dots, r$ . Se nota por  $n_j$  para  $j = 1, 2, \dots, r$  al número de unidades que no les ha ocurrido el evento justo antes del tiempo  $t_{(j)}$ , incluyendo aquellas que están próximas al evento en este tiempo. El número de eventos a este tiempo es denotado por  $d_j$ .

El mismo autor define el estimador de Kaplan-Meier para la función de sobrevivencia

$$\tilde{S}_{KM}(t) = \prod_{j=1}^k \left( \frac{n_{(j)} - d_{(j)}}{n_{(j)}} \right) = \prod_{j=1}^k \left( 1 - \frac{d_{(j)}}{n_{(j)}} \right) = \prod_{j=1}^k (1 - q_{(j)}) \quad (11)$$

Con  $q_{(j)}$  es la probabilidad de que a una unidad le ocurra el evento en el periodo  $[t_{(j-1)}; t_{(j)})$  dado que antes del tiempo  $t_{(j-1)}$  no le ha ocurrido; entonces se obtiene la expresión:

$$q_{(j)} = [t_{(j-1)}; t_{(j)}) | T \geq t_{(j-1)}]$$

Los intervalos de confianza según Diaz (2015), se definen de la siguiente forma: Un intervalo del  $100 \times (1 - \alpha)\%$  de confianza para  $S(t)$ , para un valor específico  $t$ , es dado por:

$$\tilde{S}_{KM}(t) \pm Z_{(1-\frac{\alpha}{2})} \sqrt{\widehat{var} \tilde{S}_{KM}(t)} \quad (12)$$

### 2.3.3.2. Comparación de funciones de Sobrevivencia

Como uno de los propósitos del proyecto es construir las funciones de sobrevivencia respecto a la deserción, para cada uno de los programas del estudio, resulta útil el poder hacer la comparación, la cual se puede hacer de manera visual, exploratoria o usando la estadística de log-rank.

Esta estadística es apropiada de usar cuando la razón de las funciones de riesgo de los grupos a compararse es aproximadamente constante, ya que la estadística log-rank se explica como la diferencia entre el número observado de eventos (fallas) en cada grupo y una cantidad que, para varios propósitos, puede ser considerada como el correspondiente número esperado de eventos (fallas) bajo la hipótesis nula; es decir que no hay diferencia entre los eventos observados sobre las unidades en los dos grupos. (Díaz, 2015)

Supongánsé que al momento  $t_j$  ocurren  $d_{(j)}$  eventos y  $n_{(j)}$  unidades expuestas al evento (en riesgo) en un tiempo inmediatamente inferior a  $t_j$  de la muestra combinada y, también,  $d_{jg}$  y  $n_{jg}$  en la muestra  $g = 1, 2$  y  $j = 1, \dots, k$ . En cada tiempo al evento  $t_j$ , los datos pueden ser expuestos en una tabla de contingencia  $2 \times 2$ , de la siguiente manera:

Tabla 1. Comparación de curvas de sobrevida

Evento	Grupos		Total
	1	2	
Ocorre	$d_{j1}$	$d_{j2}$	$d_j$
No Ocorre	$n_{j1} - d_{j1}$	$n_{j2} - d_{j2}$	$n_j - d_j$
Total	$n_{j1}$	$n_{j2}$	$n_j$

Fuente: Díaz (2015)

La utilidad de una estadística aproximada es que puede verificar la igualdad de las dos funciones de sobrevida mediante la ecuación:

$$LR = \frac{\sum_{j=1}^k [d_{j2} - \mu_{j2}]^2}{\sum_{j=1}^k \sigma^2 d_{j2}} \quad (13)$$

Esta expresión para muestras de tamaño grande, tiene distribución Ji-cuadrado con 1 grado de libertad.

### 2.3.3.3. Estimación de función de Riesgo acumulado

El llamado estimador de Nelson-Aalen. Dado que la función de supervivencia se puede expresar en términos de la función de riesgo,  $H(t) = -\ln S(t)$ , si se desea estimar la función de riesgo acumulada, entonces se tiene que:

$$\hat{H}(t) = -\ln \hat{S}(t) \quad (14)$$

Donde  $\hat{S}(t)$  es el estimador de Kaplan-Meier de la función de sobrevida.

También es importante en el proyecto determinar las variables que afectan el tiempo de falla, tiempo hasta la ocurrencia del evento (deserción y graduación) en cada programa que hacen parte del estudio, por ello a continuación se presenta el modelo de regresión de Cox.

## 2.4. MODELO DE RIESGOS PROPORCIONALES

Esta división del capítulo presenta a los modelos de sobrevivencia como modelos que involucran covariables al incorporar la manera en que estas afectan el tiempo de falla del sujeto en estudio, uno de estos es el denominado modelo de riesgos proporcionales, el cual es el más usado en bioestadística y en otras áreas, lo cual se debe al fácil entendimiento e implementación, además de los resultados que este proporciona al ser utilizado adecuadamente (García, 2014).

El modelo descrito en el párrafo anterior también es llamado modelo de regresión de Cox, que permite estimar la relación que hay entre un conjunto de variables explicativas fijas  $X_1, X_2, \dots, X_n$ , también conocidas como covariables, y la respuesta o tiempo de sobrevida.

La función de tasa de riesgo del tiempo de falla de un modelo con vector de covariables dadas por  $X$  está definida de esta manera:

$$h(t; X) = h_0(t) \exp(\beta^T X) \quad (15)$$

Donde  $h_0(t)$  es la función de riesgo base, sin ningún parámetro o covariable; y  $\beta^T = (\beta_1, \beta_2, \dots, \beta_n)$  es el vector de parámetros de regresión.

El modelo de Cox se dice que es un modelo semi - paramétrico, debido a que incluye una parte paramétrica y otra parte no paramétrica.

- i) La parte paramétrica,  $\exp(\beta^T X)$  llamada *función de riesgo relativo*, la cual está claramente especificada y describe los efectos relativos de los parámetros de regresión estimados sobre el riesgo.
- ii) La parte no paramétrica es  $h_0(t)$ , la función de riesgo base, que es una función arbitraria y no especificada.

El modelo de regresión de Cox tiene el cociente expresado en la siguiente fórmula conocida como *razón de riesgos relativos*, este cociente se refiere al caso entre el riesgo para dos sujetos con el mismo vector de covariables es constante en el tiempo, es decir:

$$\frac{h(t; X_i)}{h(t; X_j)} = \frac{h_0(t) \exp(\beta^T X_i)}{h_0(t) \exp(\beta^T X_j)} = \frac{\exp(\beta^T X_i)}{\exp(\beta^T X_j)} = \exp(\beta^T (X_i - X_j))$$

### 2.4.1. Estimación de parámetros del modelo de COX

El proceso usado en la estimación de los parámetros de un modelo de regresión es la verosimilitud que consiste en estimar los parámetros que maximicen el logaritmo natural de la verosimilitud, en el modelo de Cox, este procedimiento no funciona debido al desconocimiento de la función de riesgo base que aparece en el modelo de Cox.

Según Colosimo & Ruiz Giolo (2006), la presencia del componente no paramétrico  $\lambda_0(t)$  en la función de verosimilitud, lo hace inapropiado, una solución para ello es condicionar la verosimilitud para eliminar la perturbación existente, lo cual Cox propuso en un artículo en 1975, denominando al método “*función de verosimilitud parcial*” que depende solamente de los parámetros de interés.

Cox propuso la siguiente función de verosimilitud para estimar  $\beta$ :

$$L(\beta) = \prod_{i=1}^n \left[ \frac{\exp \beta^T X_i}{\sum_{j \in R(t_i)} \exp \beta^T X_j} \right]^{\delta_i} \quad (16)$$

Siendo  $\delta_i$  el indicador de censura de la siguiente forma,  $\delta_i = 0$  si la observación es censurada o, por el contrario  $\delta_i = 1$ .

Además  $R(t_i) = \{j : t_j > t_i\}$  es el conjunto que contiene los sujetos en riesgo en el tiempo  $t_i$  es decir, que están en observación y no han presentado el evento.

Puesto, que  $\delta_i$  es el indicador de censura y será 0 cuando la observación esté censurada, la función de verosimilitud parcial quedará:

$$L(\beta) = \prod_{i=1}^m \frac{\exp \beta^T X_i}{\sum_{j \in R(t_i)} \exp \beta^T X_j} \quad (17)$$

Donde  $i$  numera los  $m$  tiempos de eventos observados,  $t_1, t_2, \dots, t_m$ .

La estimación de los parámetros  $\beta$  se obtiene maximizando la función anterior de verosimilitud parcial, sin necesidad de estimar la función de riesgo base  $h_0(t)$ . Para los datos con tiempos observados empatados, se usan aproximaciones para la función de verosimilitud parcial. Una aproximación sugerida por Breslow (1974),

considera que los  $d_i$  eventos al tiempo  $t_i$  son distintos y ocurren secuencialmente que vienen dados por la siguiente fórmula:

$$L(\beta) = \prod_{i=1}^m \frac{\exp \beta^T X_i}{\left[ \sum_{j \in R(t_i)} \exp \beta^T X_j \right]^{\delta_i}}$$

Al determinar los estimadores de  $\beta$ ,  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_n)$  se obtiene el estimador de  $h_j$  la función de riesgo inicial entre  $t_{j-1}$  y  $t_j(t_0 = 0)$  con  $j = 1, 2, \dots, d$ , es decir:

$$\hat{h}_j = \frac{1}{t_j - t_{j-1}} \frac{1}{\sum_{k \in R(t_j)} \exp(\beta^T X_k)} \quad (18)$$

Después de ajustar el modelo de Cox, se verifica que sean significativas las variables del modelo, usando las tres pruebas estadísticas aplicadas en la práctica: La prueba de la razón de verosimilitud, la prueba de Wald y la prueba de puntajes, que se utilizan para juzgar la hipótesis de que los parámetros del modelo de Cox son significativos. (Barrera, 2008)

En ellas, se considera el vector de parámetros estimados  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_n)^T$  y la matriz de información evaluada en  $\beta$ , definida como:

$$I(\beta) = \frac{\partial^2 \ln L(\beta)}{\partial \beta_k^2}, k = 1, \dots, n.$$

La prueba de la razón de verosimilitudes: Denotada por  $G$ , se calcula como dos veces la diferencia entre el logaritmo natural de la verosimilitud parcial del modelo que contiene todas las variables y el logaritmo de la verosimilitud parcial del modelo sin las variables, es decir:

$$G = 2\{L_p(\hat{\beta}) - L_p(0)\} \text{ donde, } L_p(0) = -\sum_{i=1}^k \ln(n_i)$$

Con  $L_p(0)$  representa la verosimilitud del modelo con cero variables y  $n_i$  como el número de individuos en riesgo en el tiempo de supervivencia observado  $t_i$ . Bajo la hipótesis nula que los coeficientes son iguales a cero, el estadístico  $G$  se distribuye chi-cuadrado con un grado de libertad por cada variable en el modelo.



Los estadísticos Wald y puntajes: Para multivariado usan cálculos matriciales. Se denota el vector de primeras derivadas parciales de la función parcial de verosimilitud evaluado en 0 como  $U(0) = U(\beta)/_{\beta=0}$ , con la hipótesis nula de todos los coeficientes son iguales a cero y otras condiciones de la misma función, el vector se distribuye normal multivariado con media cero y matriz de covarianzas dada por la matriz de información evaluada en el vector 0,  $I(0) = I(\beta)/_{\beta=0}$ .

La prueba Wald: Hace referencia a la razón entre el coeficiente estimado y la desviación estándar de ese coeficiente como se muestra para el caso univariado:

$$Z = \frac{\hat{\beta}}{\widehat{SE}(\hat{\beta})}$$

Teniendo la hipótesis nula del coeficiente igual a cero, con este estadístico se distribuye normal estándar.

Para el caso multivariado se obtiene del hecho de que  $\hat{\beta}$ , el estimador de los coeficientes se distribuirá asintóticamente normal multivariado con vector de medias igual a cero y matriz de covarianzas dadas por la ecuación de la varianza. Entonces el estadístico de Wald se escribe así:

$$\hat{\beta}' I(\hat{\beta}) \hat{\beta}$$

Por lo cual la hipótesis nula de que todos los coeficientes son iguales a cero la expresión anterior se distribuye chi-cuadrado con  $p$  grados de libertad.

La prueba de puntajes: Es la razón de la derivada del logaritmo natural de la verosimilitud parcial y la raíz cuadrada de la información observada evaluadas las dos en  $\beta = 0$ , así en el caso univariado:

$$\hat{Z} = \frac{\partial L_p / \partial \beta}{\sqrt{I(\beta)}}, \quad \beta = 0$$

Con la hipótesis nula del coeficiente igual a cero, estadístico con distribución normal estándar. En el caso multivariado se calcula como:  $U'(0) = [I(0)]^{-1}(0)$

Bajo la hipótesis nula que todos los coeficientes son iguales a cero la expresión anterior se distribuye chi-cuadrado con  $p$  grados de libertad.

### Interpretación del modelo

La interpretación de los parámetros correspondientes a los diferentes tipos de variables explicativas del modelo de riesgos proporcionales depende si éstas son variables o factores.

Si el coeficiente de regresión ( $\beta$ ) corresponde a una variable, la cantidad  $\exp(\beta)$  es el cambio en la función de riesgo por cada unidad que se incremente la variable, dado que las otras covariables están fijas, esto se puede escribir como sigue:

$$\begin{aligned} & \frac{h(t, x_1, \dots, x_i + 1, \dots, \beta_1, \dots, \beta_p)}{h(t, x_1, \dots, x_i, \dots, \beta_1, \dots, \beta_p)} \\ &= \frac{h_0(t) \exp \{ \beta_1 x_1 + \dots + \beta_i (x_i + 1) + \dots + \beta_p x_p \}}{h_0(t) \exp \{ \beta_1 x_1 + \dots + \beta_i x_i + \dots + \beta_p x_p \}} \\ &= \exp \{ \beta_i \}. \end{aligned}$$

Si el coeficiente de regresión  $\beta$  corresponde a un factor, la interpretación es de la siguiente manera: suponga que se tiene un factor  $A$  con  $a$  niveles, y sean  $\beta_2, \dots, \beta_a$  los coeficientes de regresión de los correspondientes niveles, teniendo en cuenta el primer nivel del factor  $A$  como cero. La cantidad  $\exp\{\beta_i - \beta_j\}$  es el cociente de la función de riesgo para el sujeto con nivel  $i$  y nivel  $j$  de las variables explicativas ( $i, j = 2, \dots, a$ ), dado que las otras variables explicativas toman valores iguales. La cantidad  $\exp\{\beta_i\}$  corresponde al cambio relativo en la función de riesgo para los sujetos con las variables explicativas con nivel  $i = (i=2, \dots, a)$  y nivel 1.

#### 2.4.2. Selección del modelo

Para determinar cuál es el modelo óptimo existen diferentes métodos de selección de variables, los cuales son:

i. Forward (hacia adelante): en este método se inicia el proceso con un modelo nulo, solo con el término independiente, se ajusta un modelo con el método de máxima verosimilitud y se calcula el estadístico Chi-cuadrado con el p-valor de incluir cada variable por separado. Luego, se selecciona el modelo con la variable más

significativa, es decir, que tiene un p-valor ( $p < 0,05$ ). De nuevo, se ajusta un modelo con las variables seleccionadas y se calcula el p-valor resultante de añadir cada una de las variables que no han sido seleccionadas por separado. Tras escoger el modelo con la variable más significativa, se repiten estos pasos hasta que no queden variables significativas por incluir.

ii. Backward (hacia atrás): consiste en empezar con un modelo que contiene todas las variables candidatas e ir eliminando, una a una, cada covariable, a la vez que se calcula la pérdida de ajuste al eliminarla. Se omite del modelo la variable menos significativa, esto es, aquella que mayor p-valor tenga y se repiten estos pasos hasta que todas las variables incluidas sean significativas y no pueda eliminarse ninguna sin que se pierda ajuste de acuerdo al indicador seleccionado.

iii. Stepwise: este método es una combinación de los procesos forward y backward. Se puede comenzar, o bien, con el modelo nulo, o bien, con el modelo completo, pero en cada paso se exploran las variables incluidas por si deben salir y las no seleccionadas por si deben entrar en el modelo. Se repiten estos pasos hasta que todas las variables incluidas sean significativas y no entre ni salga ninguna más.

En este trabajo, se utiliza el método forward (hacia adelante) y el criterio de AIC (Akaike's Information Criterion) que es mencionado a continuación:

Criterio AIC: El criterio de información de Akaike (AIC) es una medida de la calidad relativa de un modelo estadístico, para un conjunto dado de datos. Como tal, el AIC proporciona un medio para la selección del modelo. AIC maneja un trade-off (compensación) entre la bondad de ajuste del modelo y la complejidad del modelo. Se da una estimación relativa de la información pérdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos. En general AIC es:  $AIC = 2k - 2 \ln(L)$

Donde  $k$  es el número de parámetros en el modelo estadístico y  $L$  es el máximo valor de la función de verosimilitud para el modelo estimado.

#### **2.4.3. Evaluación del modelo**

En el proceso de evaluación del modelo de Cox, se cuenta con un análisis de los residuos, es decir que un residuo es el valor calculado, para cada individuo, como la diferencia entre el valor de supervivencia observado y el valor estimado por la

ecuación de regresión, cuanto mayor es esa diferencia mayor será el valor del residuo, con su signo correspondiente.

El análisis de residuos en cualquier modelo estadístico según García (2014), verificar la adecuación del modelo ajustado por medio de inspección de gráficos. De manera que los residuos en el modelo de Cox pueden ser utilizados para:

- Descubrir la forma funcional correcta de un predictor continuo.
- Identificar los sujetos que están pobremente predichos por el modelo.
- Verificar el supuesto del modelo de regresión de Cox.

Dentro de este hay seis tipos de residuos: residuos de Cox-Snell, de martingala, de desvíos (deviance), de puntajes (score) y de Schoenfeld. De estos residuos pueden derivarse otros, como los dfbetas, a continuación se presentan aquellos de interés para abordar este estudio:

I. Residuos de Cox-Snell: Este tipo de residuos, desarrollados por Cox & Snell, sirven para evaluar el ajuste global del modelo planteado. Si el modelo de regresión de Cox definido por la ecuación [15] es adecuado, entonces las estimaciones del tiempo de sobrevida del modelo planteado vienen dadas por un estimador de la función de supervivencia  $\hat{S}_i(t)$ , el cual debe ser muy similar al valor verdadero de  $S_i(t)$ . Para evaluar esto, se calculan los residuos para los  $n$  individuos en estudio, de modo que:

$$r_{csi} = \widehat{H}_0(t_i) \exp(\hat{\beta}^T X_i), i = 1, \dots, n.$$

En donde  $\widehat{H}_0(t_i)$  es el estimador de la función de riesgo acumulado se define por:

$$\widehat{H}_0(s) = \int_0^s \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s) \exp(\hat{\beta}^T X_i(s))}.$$

El cual se basa en un proceso de conteo  $N_i \equiv \{N_i(t), t \geq 0\}$ , que para el  $i$ -ésimo sujeto es el número de eventos observados hasta el tiempo  $t$  y donde  $Y_i(s)$  son ceros o unos que indican si el  $i$ -ésimo sujeto está en riesgo en el tiempo  $t$ .

Para probar si los residuos de Cox-Snell están o no aproximadamente distribuidos de forma exponencial, se construye su gráfico de residuos y si estos residuos están

distribuidos de forma exponencial, entonces una estimación de la tasa de riesgo basada en  $r_{cs_i}$  representada frente a los residuos  $r_{cs_i}$  debería tender a una línea recta que pasa por el origen con pendiente igual a la unidad. Es decir, el riesgo acumulado  $H_r(r_{cs_i})$  frente a los residuos  $r_{cs_i}$ , debería ser aproximadamente una línea recta que pasa por el origen con pendiente igual a 1.

II. Residuos de Schoenfeld: estos residuos se definen como la matriz:

$$S_{ij}(\beta) = X_{ij}(t_i) - \bar{X}_j(\beta_i, t_i)$$

La cual contiene una fila por evento ocurrido y una columna por covariable, donde  $i$  y  $t_i$  son los individuos y el tiempo de ocurrencia del evento, respectivamente.

Estos residuos son útiles para verificar el supuesto de riesgos proporcionales en el modelo de Cox, pues, representados frente al tiempo observado de supervivencia deben repartirse aleatoriamente alrededor de cero, siempre que el modelo de Cox sea correcto. El estudio también puede ser abordado numéricamente. Como alternativa a la hipótesis de riesgos proporcionales, Therneau y Gramsch (2000) consideran que los coeficientes de regresión están dados por funciones de tiempo de la forma  $\beta(t) = \beta + \theta g(t)$ . En donde se contrasta la hipótesis nula:  $H_0: \theta = 0$ , el modelo de riesgos proporcionales es correcto si no se rechaza la  $H_0$ .

III. Residuos dfbeta: Sirven para determinar la influencia de cada observación en la estimación de los coeficientes de regresión. Este residuo tiene como objetivo calcular el cambio aproximado en el  $j$ -ésimo coeficiente (es decir, la  $j$ -ésima covariable) si la observación  $i$ -ésima se elimina del conjunto de datos y se vuelve a estimar el modelo sin esta observación. Así para el individuo  $i$  el valor dfbeta correspondiente a la variable  $j$  es el siguiente:

$$dfbeta_i = \beta_i - \beta_j(\text{excluyendo } i).$$

Los valores dfbeta pueden estandarizarse dividiendo por el error estándar del coeficiente correspondiente. En su representación gráfica se suelen mostrar los valores de los residuos dfbeta estandarizados para cada covariable del modelo frente a los índices de individuo (número de orden). Si la supresión de una observación hace que el coeficiente incremente, el residuo dfbeta es negativo y viceversa.

### **3. DISEÑO METODOLÓGICO**

En este capítulo se describe el proceso desarrollado para la recolección de la información, el cuál fue efectuado con los estudiantes de los programas académicos que conforman el estudio.

#### **3.1. DISEÑO MUESTRAL**

Se presenta el proceso para la conformación del diseño muestral utilizado, se estableció la población, el tipo de muestreo, marco muestral, unidades de muestreo, error de muestreo, tamaño y selección de la muestra.

##### **POBLACIÓN**

Está conformada por los estudiantes que ingresaron a los programas de estudio, Diseño Industrial e Ingeniería Electromecánica, tomando 10 cohortes así, desde primer semestre de 2004 hasta segundo semestre 2009, lo anterior debido a que tal selección en el tiempo permite la trazabilidad completa a 2015 de un estudiante que ingresó a la universidad en 2009 y desde ahí se escogieron 5 años antes, tal como lo sugieren los procesos de autoevaluación.

Según la información suministrada por la oficina de Control y Registro Académico, la población entre ambos programas académicos en estudio está conformada por 999 estudiantes. Considerando el gran tamaño de esta población en cada programa y con el fin de minimizar los costos, conseguir una mayor rapidez y exactitud, se determinó la selección de una muestra representativa aleatoria sin reemplazo, además como menciona Pérez (2000) al aproximar las características poblacionales mediante estimadores basados en la muestra se comete un error, error que mide la representatividad de dicha muestra, por lo cual se fija un error de muestreo mínimo para de esa forma determinar el tamaño de muestra.

##### **TIPO DE MUESTREO**

Se utilizó un Muestreo Aleatorio Simple sin Reposición (M.A.S.) para seleccionar la muestra en cada uno de los programas a estudiar, en este muestreo se realiza la selección de un subconjunto de elementos de la población en forma aleatoria sin reposición, los elementos tienen la misma probabilidad de ser seleccionados y el orden no interviene, como el muestreo es con probabilidades iguales, todas las muestras son equiprobables, por ello, el M.A.S es un método de selección de muestras en las cuales las unidades se eligen individual y directamente por medio de un proceso aleatorio en el que cada unidad no seleccionada tiene la misma

oportunidad de ser elegida que todas las otras unidades en cada extracción de la muestra. (Lininger & Warwick, 1978)

## UNIDAD DE MUESTREO Y MARCO MUESTRAL

La unidad elemental de muestreo corresponde al estudiante de la UPTC – Duitama que pertenece a uno de los programas estudiados en este proyecto, durante el periodo de tiempo de cinco años que corresponde a diez semestres calendario desde el primer semestre de 2004 hasta el segundo semestre de 2009.

Cuyo marco muestral se construyó de acuerdo con la información suministrada por el Sistema para la prevención de la deserción en las instituciones de Educación superior -SPADIES- y la oficina de control y registro académico -SIRA- el marco muestral coincide con la población del estudio, el programa Diseño Industrial posee 461 estudiantes, Ingeniería Electromecánica 538 estudiantes para un marco muestral total de 999 estudiantes.

## TAMAÑO Y SELECCIÓN DE LA MUESTRA

A continuación se tiene en primer lugar el proceso para hallar el tamaño de la muestra en cada programa académico perteneciente al estudio y después se determinar la selección de la muestra realizada con el muestreo aleatorio simple sin reposición a través del programa R.

Los parámetros para calcular los tamaños de cada muestra son:

Tabla 2. Parámetros para determinar el tamaño de la muestra

Parámetro	Descripción	Programa Académico	
		Diseño industrial (DI)	Ingeniería Electromecánica (ING)
<b>n</b>	El tamaño de la muestra que se desea calcular	170	179
<b>N</b>	Total de estudiantes matriculados, de cada programa de administración de la UPTC Duitama.	461	538
<b>e</b>	Error máximo aceptable, $e = 0.06$ ó 6%		
<b>P</b>	Es la proporción que esperamos encontrar= 0.5		
<b>1 - p</b>	0.5		
<b>Z</b>	Valor correspondiente de la curva normal con una confiabilidad del 95%=1,96		

Fuente: La Autora

Para la estimación se usará una proporción, se estima el tamaño de la muestra según el supuesto finito de la población, a partir de la siguiente fórmula:

$$n = \frac{N \cdot Z^2 \cdot (p) \cdot (1 - p)}{(N - 1) \cdot e^2 + Z^2 \cdot (p) \cdot (1 - p)}$$

✓ **Muestra para el programa de Diseño Industrial (DI)**

$$n = \frac{(461) \cdot (1,96)^2 \cdot (0,5) \cdot (1 - 0,5)}{(461) \cdot (0,06)^2 + (1,96)^2 \cdot (0,5) \cdot (1 - 0,5)}$$

$$= 168,986 \approx 170$$

✓ **Muestra para el programa de Ingeniería Electromecánica (ING)**

$$n = \frac{(538) \cdot (1,96)^2 \cdot (0,5) \cdot (1 - 0,5)}{(538) \cdot (0,06)^2 + (1,96)^2 \cdot (0,5) \cdot (1 - 0,5)}$$

$$= 178,564 \approx 179$$

Usando el Software estadístico R, se seleccionó de forma aleatoria el listado de estudiantes pertenecientes a los programas de Ingeniería Electromecánica y Diseño Industrial de la Universidad Pedagógica y Tecnológica de Colombia de Duitama, que conformarán la muestra del estudio, el listado de estos estudiantes se muestra en los Anexos A y B, respectivamente.

### 3.2. RECOLECCIÓN DE LOS DATOS

La recolección de los datos se refiere a la descripción de las fuentes de información usadas para recoger las mediciones de las variables objeto de estudio en los estudiantes de la muestra.

Inicialmente se acudió a los registros que suministra el SPADIES (Sistema para la prevención de la deserción en las instituciones de Educación superior) y se complementó la medición con la información brindada por la Oficina de Registro y Control Académico de Tunja, particularmente la base de datos SIRA (Sistema de información y Registro académico), en los siguientes párrafos se describe en orden los sistemas usados SIRA y SPADIES, se aclara que en algunos casos no se logró completar la base de datos para todos los estudiantes.



### **Sistema para la Prevención y Análisis de la Deserción en las Instituciones de Educación Superior (SPADIES)**

Este sistema fue propuesto en el año 2002 por el MEN, creado con el fin de disminuir la deserción universitaria, ser una herramienta que obtuviera la información por sectores y a la vez el total de la deserción. Actualmente el SPADIES está instalado en el 99% de las instituciones de educación superior, este sistema hace seguimiento a más de 3 millones de estudiantes; esa información se encuentra disponible al público de dos formas, una es la consulta on-line y la otra es mediante un software de aplicación.

Para este proyecto se usó el software de aplicación del Sistema para la Prevención y Análisis de la Deserción en las Instituciones de Educación Superior (SPADIES) (Actualizado hasta Abril 09 de 2015) que posee los datos de los estudiantes de la Universidad Pedagógica y Tecnológica de Colombia, señalando la sede a que pertenecen las carreras que se estudian, para este caso Seccional Duitama, la búsqueda se realizó por individuo, apoyándose en un dato personal como lo es el código que recibe el estudiante al ingresar a la universidad, de esa forma se hallaron estas variables: cohorte, nombres, apellidos, documento, fecha de nacimiento, sexo, programa, código de estudiante, puntaje (estandarizado) examen de estado, nivel educativo de la madre, número de hermanos, posición entre hermanos, edad de presentación del examen de estado, vivienda propia, trabaja al momento de presentar el examen de estado, ingreso familiar al presentar examen de estado, estrato, nivel de SISBEN, personas hogar, estado académico y período en que se graduó o desertó.

Además se halla una gráfica que se genera cuando seleccionamos un registro de un estudiante, la línea roja está debajo de la línea azul nos indica que el alumno está en peligro de deserción. La supervivencia estructural es la que predice el modelo (azul), la otra es la que se ha observado con el estudiante (roja). En conclusión la azul es la pronosticada si todo sale bien (no pierde materias) y no recibe apoyos y la roja indica lo que realmente ocurrió durante el tiempo de estudio.

Con el fin de completar la base de datos de este proyecto se recurrió al Sistema de Información de Registro Académico (SIRA), con esta herramienta así se consolidó las siguientes variables: matriculados en la cohorte, número de cancelaciones de semestre, número de materias perdidas y número de repetición de materias. A continuación se explica más detalladamente el Sistema de Información de Registro Académico (SIRA) como sistema de apoyo.

### **Sistema de Información de Registro Académico (SIRA)**

Es el sistema académico que maneja la Universidad Pedagógica y Tecnológica de Colombia (UPTC) contiene los siguientes módulos: aspirantes, recibos, exenciones y deudas, expedición de certificaciones, control de calificaciones, inscripción de asignaturas, horarios, graduaciones, docentes y evaluación a docentes, planes de estudio, datos estadísticos, auditoria y módulos en internet.

En el caso de los datos estadísticos, este sistema se encarga de generar informes estadísticos para la oficina de Planeación relacionados con aspirantes, deserción, matriculados, admitidos, terminaciones académicas, recaudos, valores de matrícula, etc. Este módulo de datos estadísticos fue el usado para completar la información faltante en la base de datos, aunque de algunas variables no se obtuvieron todos los registros.

### **3.3. VARIABLES OBJETO DE ESTUDIO**

Variable es una característica que tiene en común todos los elementos de un conjunto de personas o cosas, al medirla en los elementos de ese conjunto, se obtienen valores diferentes e impredecibles (Lerma, 2009). Las variables que simplemente clasifican o caracterizan sectores de la población se conocen como atributo, variable categórica o cualitativa (Ospina, 2001). Se dice que es una variable categórica nominal cuando está no establece ninguna relación de orden entre sus categorías, cuando existe esta relación de orden entre categorías se llaman variables categóricas ordinales.

Cuando las características pueden ser medidas a partir de valores que son números pero que además estos conservan sentido práctico de sus propiedades intrínsecas (cuantificar la diferencia entre dos posibles valores, sumarlos, etc.), estas características se conocen como variables cuantitativas.

Cuando una variable numérica se caracteriza por la propiedad de que para dos posibles valores de ella solamente hay un número finito de posibles valores intermedios se llama variable cuantitativa discreta. En caso que entre dos posibles valores de ella cualquier valor intermedio es también un valor posible de la variable se llama variable cuantitativa continua.

A continuación se hace la descripción de las variables elegidas como explicativas según la literatura investigada, también se menciona el objeto de medición en este proyecto, indicando el tipo de variable.

Tabla 3. Listado de variables objeto de estudio

VARIABLE	DESCRIPCIÓN	TIPO DE VARIABLE
Desertó	Abandono por Artículo 42 (no renovación de matrícula por tres semestres consecutivos), retiro definitivo del o por causas académicas contempladas en el Artículo 80 literales a, b, c, d y e del reglamento estudiantil.	Cualitativa
Edad del estudiante	Edad en años cumplidos por el estudiante al ingresar al primer semestre	Cuantitativa discreta
Estado del Estudiante	Presentadas por el SIRA y corresponde al estado del estudiante según el reglamento.	Variable cualitativa
Género	Femenino o masculino según sea el caso	Cualitativa
Ingresos Familiares	En salarios mínimos legales vigentes según el periodo de ingreso del estudiante. De acuerdo a la clasificación establecida: Bajo, medio y alto.	Cualitativa
Nivel educativo de la madre	Se refiere al máximo nivel de estudios alcanzados por la madre	Cualitativa
Número de hermanos	En el momento de presentar el examen de estado.	Cuantitativa discreta
Número de Materias perdidas	Un estudiante puede perder varias veces la misma materia	Cuantitativa discreta
Número de materias repetidas	Un estudiante puede repetir más de una vez la misma materia	Cuantitativa discreta
Puntaje estandarizado ICFES	Con valores de 0 a 100	Cuantitativa continua
Semestre en el que desertó	Semestre en el que el estudiante desertó	Cuantitativa discreta
Tiempo hasta que desertó	Número de semestres desde el inicio del programa académico	Cuantitativa Discreta
Tiempo hasta que se gradúa	Número de semestres que el estudiante permanece hasta graduarse	Cuantitativa Discreta
Tipo de Vivienda	Si posee o carece en el momento de presentar el examen de estado	Cualitativa

Fuente: Autora

### **3.4. TÉCNICAS DE ANÁLISIS DE DATOS**

El programa estadístico R es usado en el área de estadística, fue inicialmente desarrollado por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993, programa que está en constante proceso de actualización, se determinó su uso en este proyecto pues permite trabajar con una gran cantidad de datos incluyendo los casos donde hay escasez de información, la utilización de este software permite obtener información precisa y confiable del modelo de sobrevida.

La decisión de usar este paquete se sustenta en la coherencia de los métodos utilizados por el paquete con la teoría desarrollada en este trabajo. La versión de R 3.0.0 (2013-04-03) Copyright (C) 2013 The R Foundation for Statistical Computing Platform: i386-w64-mingw32/i386 (32-bit), con el paquete R-Commander en su versión 2.1-7.

Entre las bibliotecas de R que permiten llevar a cabo un análisis de supervivencia, la más utilizada y empleada en este trabajo, es “survival”. Dicha biblioteca es seleccionada por ser capaz de soportar datos que presentan censura. Además, contiene numerosas rutinas utilizadas para el desarrollo de este estudio, las cuales figuran en el Anexo C.

## 4. MODELO DE SOBREVIDA PARA DISEÑO INDUSTRIAL

Este capítulo esta subdivido en dos secciones, la primera es la descripción de las variables por medio de un análisis univariado y bivariado, en la segunda sección se encuentra la descripción de la construcción del modelo de sobrevida para el programa Diseño Industrial tanto en el fenómeno de la deserción universitaria como en la terminación de los estudios académicos, en otras palabras la graduación.

### 4.1. DESCRIPCIÓN DE LOS DATOS

A continuación se presenta una descripción de las variables de estudio, resumidas en porcentajes y algunos estadísticos como la media, la mediana, mínimo y máximo del grupo poblacional. De igual manera, se presenta el análisis bivariado desde la intención del modelamiento. La descripción mencionada se realiza a través de R-Commander. Los detalles de las sentencias en R figuran en el Anexo C.

La siguiente tabla contiene los datos de las variables con las unidades de medición del programa Diseño Industrial, de acuerdo a la muestra:

Tabla 4. Resumen descriptivo de las variables de estudio de DI

<i><b>Variable</b></i>	<i><b>Unidades de medición /codificación</b></i>	<i><b>Resumen numérico</b></i>
<i>Estado del estudiante</i>	Matriculado= MT Desertor= DE Graduado= GR Terminación Académica= TA	9 (5.29%) 102 (60%) 47 (26.25%) 12 (7.06%)
<i>Edad de Presentación del Examen de Estado</i>	En años cumplidos	Media = 17.1 Min – Max = 15 – 22 Mediana = 17
<i>Ingreso Familiar al Presentar el Examen de Estado</i>	Bajo= [0-1) y [1-2) SMMLV Medio= [2-3) y [3-5) SMMLV Alto= [10) SMMLV Sin Clasificar= SC	101 (59.41%) 46 (27.06%) 2 (1.18%) 21 (12.35%)
<i>Género</i>	Femenino= F Masculino= M	64 (37.65%) 106 (62.35%)
<i>Nivel educativo de la madre</i>	Básica Primaria= BP Básica Secundaria= BS Técnico= TEC Universitario= UN Sin Clasificar= SC	36 (21.18%) 71 (41.76%) 13 (7.65%) 30 (17.65%) 20 (11.76%)

<b>Variable</b>	<b>Unidades de medición /codificación</b>	<b>Resumen numérico</b>
<i>Número de Hermanos al momento de presentar el Examen de Estado</i>	1= Uno 2= Dos 3= Tres 4= Cuatro 5= Cinco 6= Seis Sin Clasificar= SC	36 (21.18%) 76 (44.70%) 22 (12.94%) 10 (5.88%) 4 (2.35%) 2 (1.18%) 20 (11.76%)  Media = 2.173 Min – Max = 1 – 6 Mediana = 2
<i>Número de Materias Perdidas</i>	Asignaturas pérdidas durante los semestres cursados del programa académico. Sin Clasificar= SC	Media = 7.87 Min – Max = 1 - 34 Mediana = 7 23 = (13.53%)
<i>Número de Materias Repetidas</i>	Una materia se pudo perder más de una vez Sin Clasificar= SC	Media = 7.03 Min – Max = 1 - 29 Mediana = 6 4 (2.35%)
<i>Puntaje en el Examen de Estado</i>	De 1 – 100 puntos Sin Clasificar= SC	Media = 76.35 Min – Max = 1 - 100 Mediana = 82 4 (2.35%)
<i>Trabajaba al presentar el examen de estado</i>	Si= SI No= NO Sin Clasificar= SC	6 (3.53%) 146 (85.88%) 18 (10.59%)
<i>Vivienda Propia</i>	Si= SI No= NO Sin Clasificar= SC	8 (4.7%) 130 (76.47%) 32 (18.82%)
<i>Tiempo hasta la Deserción</i>	Semestres Académicos (16 semanas)	Media = 3.412 Min – Max = 1 - 17 Mediana = 2
<i>Estado del estudiante Dicotomizado</i>	Desertor= D, No Desertor= ND	102 (60%) 68 (40%)
<i>Deserción</i>	Académica= A, No Académica= NA	43 (42.15%) 59 (57.85%)

<b>Variable</b>	<b>Unidades de medición /codificación</b>	<b>Resumen numérico</b>
<i>Tipo de deserción</i>	Precoz (Hasta primer semestre) Temprana (Entre segundo y quinto semestre) Tardía (De sexto semestre en adelante)	46 (45.1%) 38 (37.26%) 18 (17.64%)
<i>Motivo de deserción Académica (Perdió cupo Artículo 80)</i>	Literal B (Promedio aritmético acumulado inferior a tres cero (3.0) o promedio aritmético semestral inferior a dos cero (2.0))= LB  Literal C (Pérdida de una asignatura que curse como repitente siendo su promedio acumulado inferior a tres cero (3.0). Con el promedio aritmético acumulado igual o superior a tres cero (3.0), la podrá cursar por tercera y última vez)= LC  Literal D (Pérdida en un mismo periodo académico de dos asignaturas que se cursen como repitente)= LD  Literal E (Quien pierda una asignatura por tercera vez)= LE	15 (34.88%)  15 (34.88%)  6 (13.95%)  7 (16.28%)
<i>Motivo de deserción no Académica</i>	Estudiantes que pasaron carta de retiro definitivo= RD  Perdió cupo Artículo 42 Literal E (Quien durante tres semestres no renueve matrícula)= NRM	38 (64.4%)  21 (35.6%)

Fuente: La Autora

De acuerdo con la tabla anterior se afirma que la mayoría de estudiantes que ingresan al programa de Diseño Industrial son hombres (62.35%), se observa la edad promedio al momento de presentar el examen de Estado con 17 años, el puntaje promedio obtenido en el examen ICFES fue 76.35 en una escala de valoración entre 0 y 100, de estos estudiantes el 76.47% cuentan con vivienda propia y el 85.88% no trabajaban en el momento de presentar el examen de Estado para ingreso a la educación superior.

Además el 52.41% de hogares de los estudiantes tenían un ingreso entre 0 y 2 Salarios Mínimos Mensuales Legales Vigentes (SMMLV) al momento de presentar

el examen de estado, otra variable relacionada con este examen es el número de hermanos ya que el 44.7% de los estudiantes tenían dos hermanos en ese momento, analizando el nivel educativo de las madres se identifica que el 41.76% estudió hasta la básica secundaria.

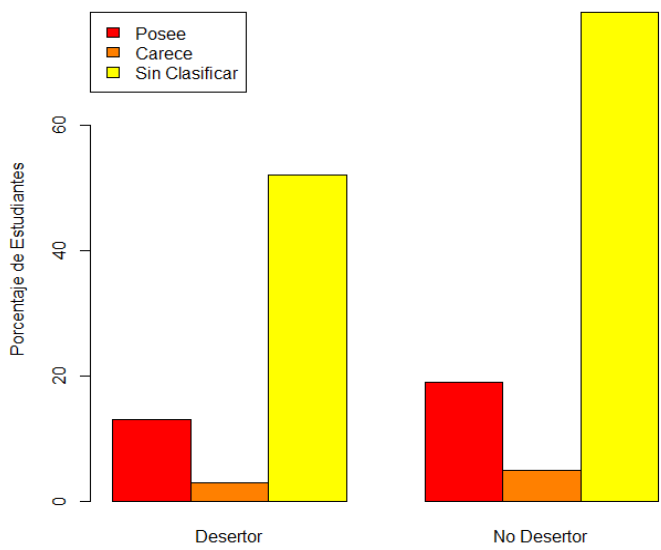
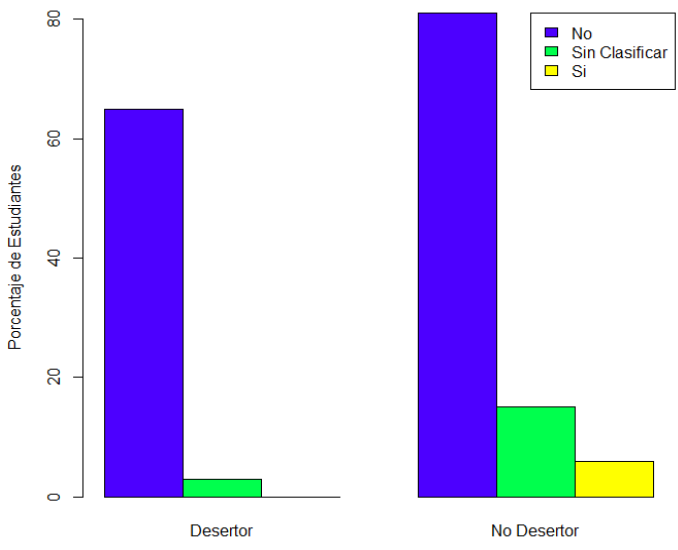
Las siguientes variables fueron tomadas hasta el primer semestre de 2015, para los estudiantes pertenecientes a la muestra se observó que el 60% de los estudiantes de las diez cohortes analizadas en este programa desertó, en los tipos de deserción sobresale que entre la deserción precoz y temprana está el 82.36% de los estudiantes desertores, eso según el caso de la deserción respecto al tiempo, en el caso de los motivos de deserción, para la académica el 69.6% de los estudiantes presentaron pérdida de cupo por los literales B y C del Artículo 80, en cuanto a la deserción no académica el 64.4% es decir la mayoría de los estudiantes desertores por este motivo pasaron la carta de retiro, y durante el mismo periodo se graduó el 26.25% de los estudiantes.

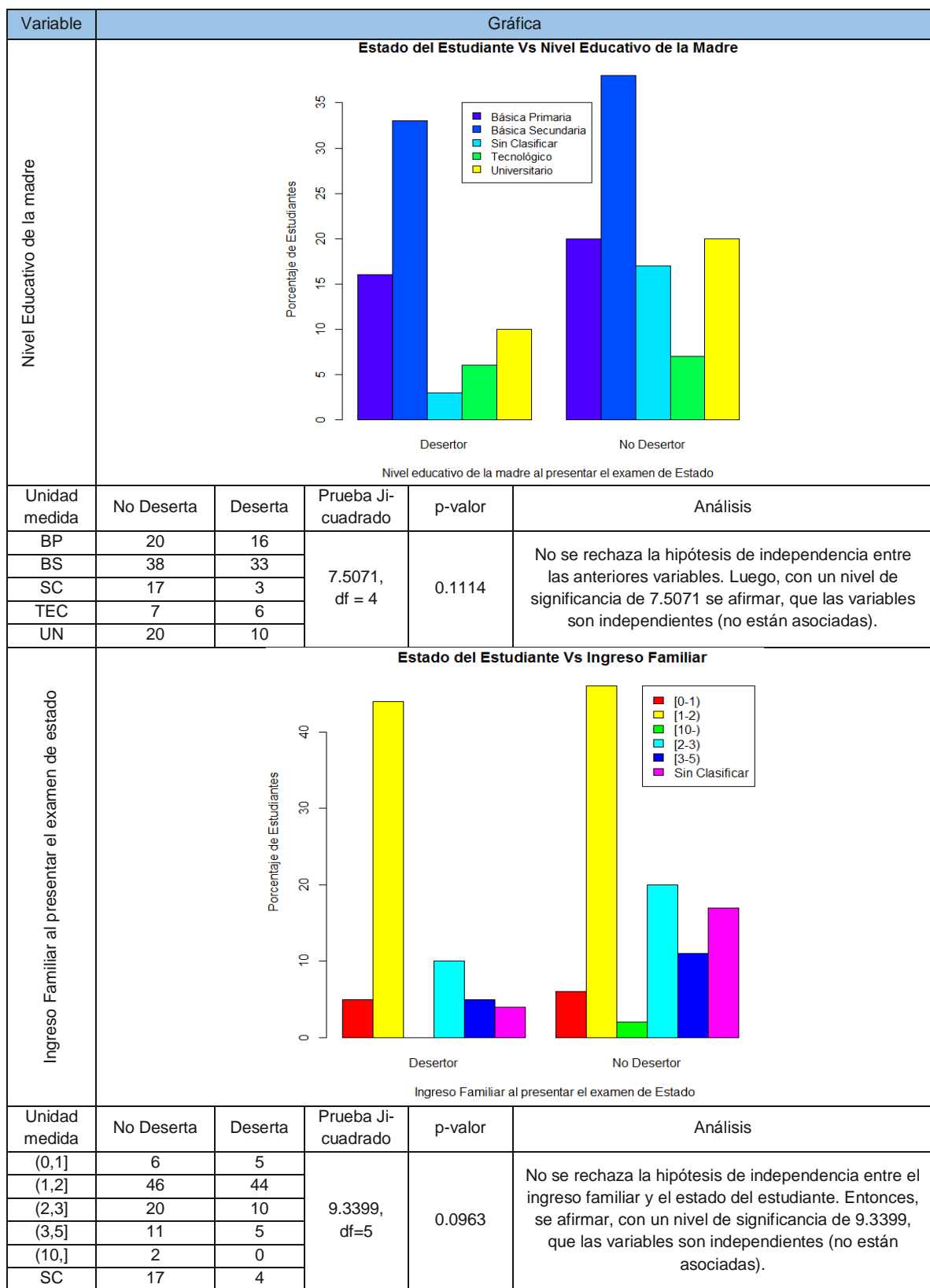
Se realiza también un análisis bivariado, con el fin de determinar si las variables cualitativas medidas en los estudiantes son independientes del estado del estudiante (desertor o no desertor). La prueba estadística usada es Ji-cuadrado.

Tabla 5. Resumen bivariado de las variables de estudio de DI

Variable	Gráfica													
Género	<div><p>Estado del Estudiante Vs Género</p><p>Porcentaje de Estudiantes</p><p>Desertor No Desertor</p><p>Estado del Estudiante</p><table border="1"><thead><tr><th>Estado del Estudiante</th><th>Femenino (%)</th><th>Masculino (%)</th></tr></thead><tbody><tr><td>Desertor</td><td>25</td><td>43</td></tr><tr><td>No Desertor</td><td>39</td><td>63</td></tr></tbody></table></div>					Estado del Estudiante	Femenino (%)	Masculino (%)	Desertor	25	43	No Desertor	39	63
	Estado del Estudiante	Femenino (%)	Masculino (%)											
Desertor	25	43												
No Desertor	39	63												
Unidad medida	No Deserta	Deserta	Prueba Ji-cuadrado	p-valor	Análisis									
M	63	43	0.0376, df = 1	0.8463	No se rechaza la hipótesis de independencia entre las variables. Luego, con un nivel de significancia de 0.0376, las variables son independientes (no están asociadas).									
F	39	25												



Variable	Gráfica				
Vivienda Propia	<div>Estado del Estudiante Vs Vivienda Propia</div>  <p>Estado de vivienda del Estudiante al presentar el examen de Estado</p>				
	Unidad medida	No Deserta	Deserta	Prueba Ji-cuadrado	p-valor
	Si	78	52	0.026, df = 2	0.9871
	NO	19	13		
	SC	5	3		
Trabajaba al Presentar el Examen de Estado	<div>Estado del Estudiante Vs Trabajaba</div>  <p>Estado de trabajo del Estudiante al presentar el examen de Estado</p>				
	Unidad medida	No Deserta	Deserta	Prueba Ji-cuadrado	p-valor
	SI	6	0	9.3265, df = 2	0.0094
	NO	81	65		
	SC	15	3		
		Análisis			
		No se rechaza la hipótesis de independencia entre las variables. Con un nivel de significancia de 0.026, las variables son independientes (no están asociadas).			



Fuente: La Autora

De acuerdo con los análisis bivariados realizados en la tabla 5, se determinó que el desertar o no, depende de si trabajaba al momento de presentar el examen de estado y del ingreso familiar.

Por último, mediante el test de Pearson se observa en la tabla 6, la relación entre el tiempo en desertar y las variables cuantitativas usadas en este estudio.

Tabla 6. Test de correlaciones de Pearson entre tiempo en desertar y variables cuantitativas en DI

Variable	Tiempo en desertar		
	Coeficiente de Correlación	Estadístico t	P
Edad ingreso al programa	0.1095	1.0795	0.283
Número de hermanos	0.1434	1.32	0.190
Puntaje en las pruebas de estado	0.1856	1.851	0.067

Fuente: La Autora

Se observa que el puntaje en las pruebas de Estado está relacionado con el tiempo de deserción. Entre mayor es el puntaje en las pruebas de estado de ingreso a la educación superior el tiempo de permanencia en el programa es mayor.

## 4.2. MODELO PARA DESERCIÓN

En este proyecto se planteó como objetivo identificar los factores que están relacionados al tiempo de la deserción de un estudiante perteneciente al programa Diseño Industrial. En primer lugar, se presentará la aplicación del método de Kaplan Meier para estimar la función de sobrevida, con ello se analiza la evolución de la probabilidad de la deserción con su respectivo intervalo de confianza, en segundo lugar, se construirá el modelo de regresión de Cox para estimar el efecto de las variables de estudio sobre los tiempos de sobrevida al evento, en este caso la deserción.

Se prosigue con el análisis del tiempo hasta la deserción, usando la información disponible, es decir los datos censurados y no censurados. Las probabilidades de sobrevivencia en cada intervalo así como la función de sobrevivencia se calculan con el estimador de Kaplan Meier, teniendo en cuenta que no se asumirá modelo probabilístico para el tiempo hasta la deserción y la existencia de datos censurados a derecha.

En la sección 2.3.1 se mencionó la definición de las características de sobrevida:

Tabla 7. Características de los datos para el modelo de sobrevivencia de DI

Unidad sobre la cual se registra el evento		Estudiante perteneciente a las cohortes entre 2004 y 2009 del programa de Diseño Industrial
El evento de interés o falla		Abandono del programa por parte del estudiante (deserción).
Variable respuesta "Tiempo hasta la deserción"		El tiempo hasta que el estudiante se va del programa ya sea por causas académicas o no académicas, la variable es discreta, ya que se mide en número de semestres cursados hasta que presenta el fenómeno de estudio.
Tiempo	Inicio del Estudio	I Semestre de 2004
	Origen del evento	Cohorte del estudiante (Primera matricula del estudiante en el programa)
	Final del Estudio	II Semestre de 2015
Censura	Tipo	Tipo I y a derecha. Los estudiantes entran al estudio en diferentes tiempos, es decir, diferentes cohortes, y el punto final del estudio es el mismo para todos. En este caso, el tiempo de censura para cada estudiante es conocido desde el momento que ingresa al estudio, con ello cada individuo tiene fijo y especificado su tiempo de censura. Se considera como censura al estudiante graduado o que continúa estudiando.
	Tiempo	Promedio de graduación del programa fue 14 semestres.

Fuente: La Autora

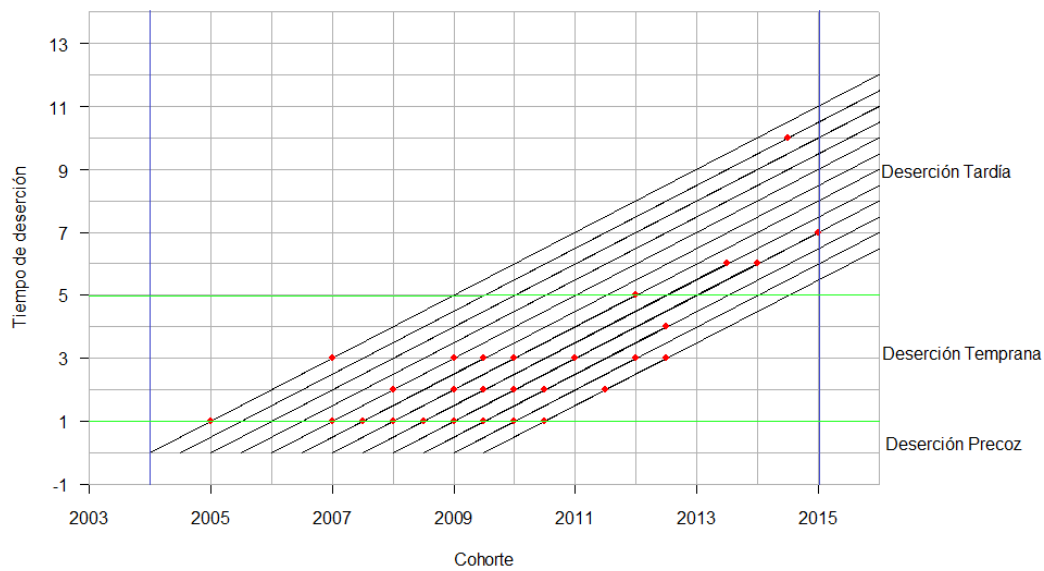
La siguiente tabla presenta el N° total que se refiere la cantidad de estudiantes pertenecientes a la muestra del programa Diseño Industrial, específicamente se estudiaron a 170 estudiantes, donde el N° de eventos indica la cantidad de desertores durante el periodo de análisis.

Tabla 8. Resumen porcentaje de censurados DI

Programa	N° Total	N° de eventos	Censurado	
			N	Porcentaje
Diseño Industrial	170	102	68	40 %

Fuente: La Autora

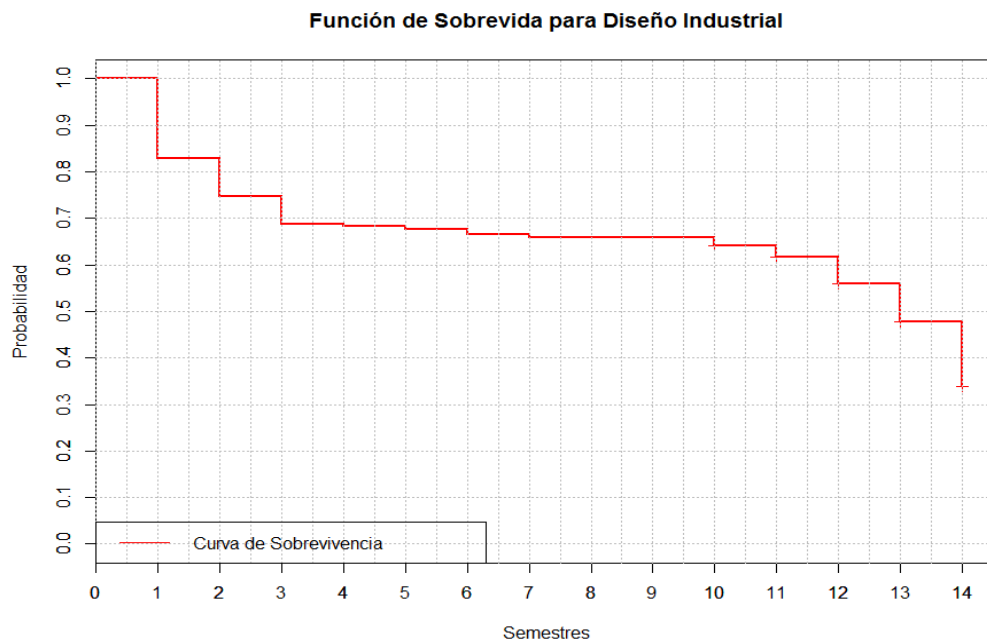
A continuación se muestra la representación del tiempo hasta la deserción de Diseño Industrial mediante el diagrama de Lexis, cuya sintaxis R aparece en el Anexo C. Diagrama que refleja en el eje horizontal el tiempo calendario y la longitud del tiempo de vida, representada por una línea a 45° el tiempo que un individuo pasa en el estudio está representado por la altura del rayo en el eje vertical.



Gráfica 1. Diagrama de Lexis para DI

En la gráfica 1 se evidencia que los estudiantes bajo estudio no tienen el mismo tiempo de origen como se mencionó en la caracterización de los datos, los puntos rojos representan aquellos estudiantes que presentaron el evento y el tiempo en que sucedió (expresado en semestres). La gráfica muestra que la mayoría de estudiantes desertan precozmente y unos pocos registran deserción tardía.

#### 4.2.1. Función de sobrevivencia



Gráfica 2. Función de sobrevivencia estimada para deserción en DI

Con el fin de realizar la estimación de la función de Sobrevida se utiliza el método de Kaplan Meier, esta se obtiene a través de R-Commander, mediante la función survfit, en el Anexo C se muestra la sintaxis de R. En el gráfico 2 y en la tabla 9 se observan las probabilidades de sobrevida en cada intervalo.

Se observa que a medida que los semestres aumentan la probabilidad de continuar en el programa disminuye. En los tres primeros semestres se presentan los mayores decrecimientos. Del cuarto al décimo semestre la probabilidad es continua en el programa se mantiene constante. Y un estudiante que llegue al décimo semestre cuando el estudiante tiene una probabilidad del 33.8% de continuar en el programa.

Tabla 9. Estimaciones función de sobrevida para deserción en DI

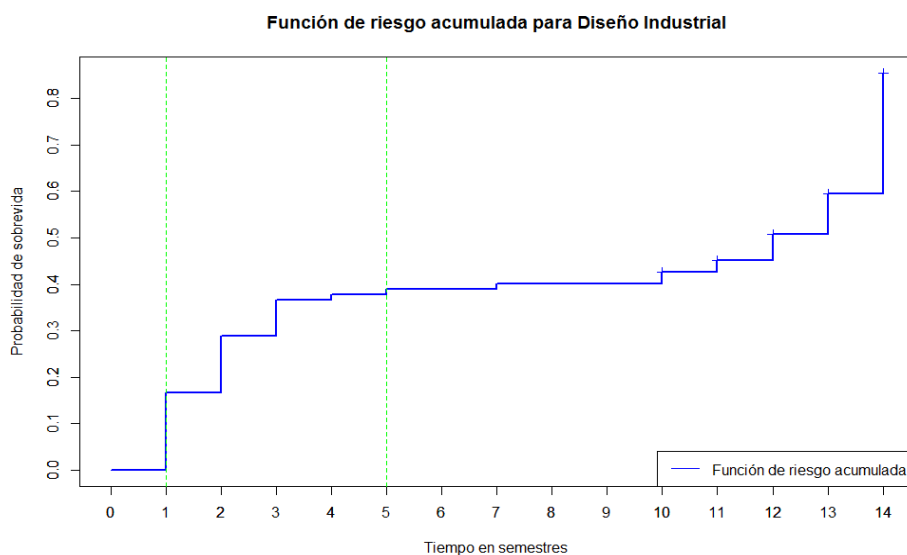
Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	170	29	0.829	0.0288	0.775	0.888
2	141	14	0.747	0.0333	0.684	0.815
3	127	10	0.688	0.0355	0.622	0.762
4	117	1	0.682	0.0357	0.616	0.756
5	116	1	0.676	0.0359	0.610	0.751
6	115	2	0.665	0.0362	0.597	0.740
7	113	1	0.659	0.0364	0.591	0.734
10	112	3	0.641	0.0368	0.573	0.717
11	108	4	0.617	0.0373	0.548	0.695
12	96	9	0.560	0.0385	0.489	0.640
13	76	11	0.479	0.0399	0.406	0.564
14	58	17	0.338	0.0402	0.268	0.427

Fuente: La Autora

La función de riesgo, también conocida como la tasa instantánea de mortalidad, describe la forma en que cambia la tasa instantánea de la deserción al paso del tiempo, es decir el semestre. La función de riesgo acumulado, permite tener la información del comportamiento del riesgo a lo largo del tiempo. A continuación se presenta la función de riesgo correspondiente a Diseño Industrial, para efectos de la interpretación se indican los tipos de deserción en la gráfica, específicamente el semestre uno (deserción precoz) y hasta el semestre número cinco (deserción temprana).

De acuerdo a la función de riesgo acumulado se puede afirmar que la probabilidad que un estudiante deserte justo al terminar el primer semestre es de 19%, es decir

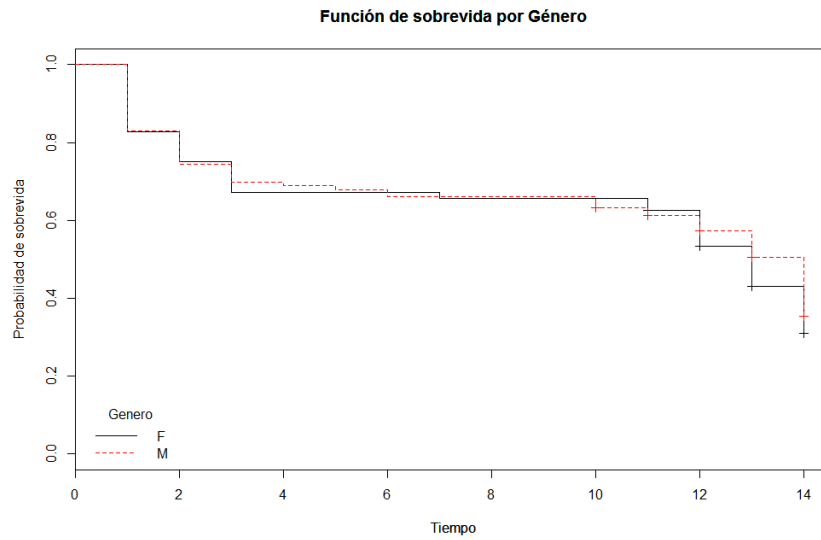
la probabilidad que deserte precozmente, de igual manera se observa que la función de riesgo presenta los mayores cambios en la deserción temprana aunque después del semestre doce vuelve a incrementarse la probabilidad de desertar, por ejemplo, se aprecia, que en tercer semestre la probabilidad de que habiendo llegado a segundo deserte en el tercer semestre es del 8.2%. En cuanto a la probabilidad de desertar al llegar al semestre catorce terminando el semestre trece es del 34.4%.



Gráfica 3. Función de riesgo acumulado para deserción de DI

El fenómeno de la deserción no se estudia de manera aislada, razón por la cual se hace necesario un análisis de la función de sobrevivencia separado por poblaciones de estudio según las variables categóricas, presentado a continuación. Primero, se tiene la función de sobrevivida de acuerdo al género.

Para analizar la función de sobrevivida por género se tiene en cuenta la gráfica 4 y la tabla 10, se evidencia en la deserción precoz que el porcentaje de sobrevivida esta entre 83% y 82% para hombres y mujeres respectivamente, por lo cual la variación es mínima. En la deserción temprana por ejemplo en el cuarto semestre se mantiene la misma situación para hombres y mujeres pues tienen un porcentaje de sobrevivida cercano al 70% y por último en la deserción tardía en el género masculino tiene un porcentaje de 66% y en el género femenino con mujeres 67.2% en el semestre número seis, con lo cual se mantiene la poca variabilidad en las categóricas de esta variable. Además, al realizar la comparación se obtuvo un valor de  $\chi^2 = 0.3$  con 1 grado de libertad y un p-valor de 0.615, se puede concluir que la sobrevivida tiene el mismo comportamiento en hombres y mujeres.



Gráfica 4. Función de sobrevivida por género en DI

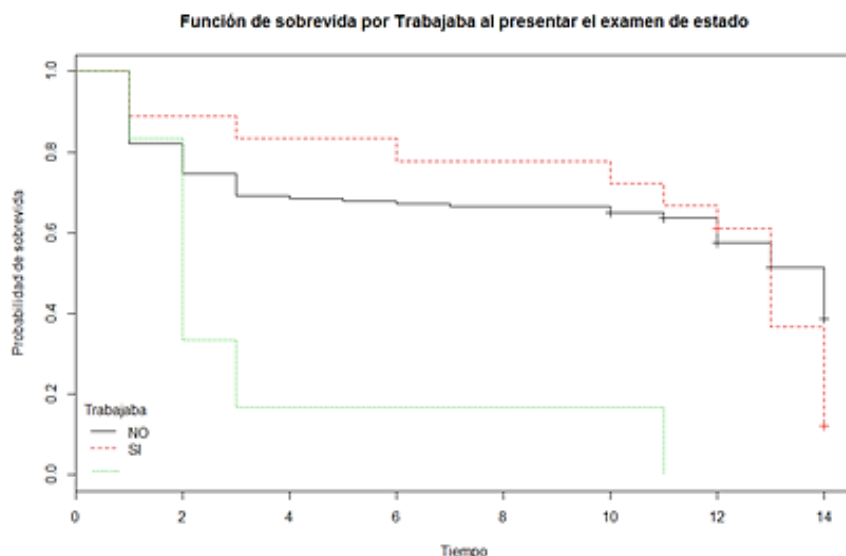
Tabla 10. Resumen estimador Kaplan Meier por género de DI

<b>Género=F</b>					<b>95% IC</b>	
Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	64	11	0.828	0.0472	0.741	0.926
2	53	5	0.750	0.0541	0.651	0.864
3	48	5	0.672	0.0587	0.566	0.797
7	43	1	0.656	0.0594	0.550	0.784
11	42	2	0.625	0.0605	0.517	0.756
12	34	5	0.533	0.0641	0.421	0.675
13	26	5	0.431	0.0662	0.319	0.582
14	18	5	0.311	0.0659	0.205	0.471
<b>Género=M</b>					<b>95% IC</b>	
Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	106	18	0.830	0.0365	0.762	0.905
2	88	9	0.745	0.0423	0.667	0.833
3	79	5	0.698	0.0446	0.616	0.791
4	74	1	0.689	0.0450	0.606	0.783
5	73	1	0.679	0.0453	0.596	0.774
6	72	2	0.660	0.0460	0.576	0.757
10	70	3	0.632	0.0468	0.547	0.731
11	66	2	0.613	0.0473	0.527	0.713
12	62	4	0.573	0.0482	0.486	0.676
13	50	6	0.505	0.0500	0.416	0.613
14	40	12	0.353	0.0506	0.267	0.468

Fuente: La Autora



También se realizó la prueba para la hipótesis nula de igualdad de las curvas de sobrevivencia respecto a si trabajaban o no a la hora de presentar el ICFES, encontrándose que, con  $\text{Chisq} = 13.3$  con 2 grados de libertad y un p valor de 0.00126, se afirma que las curvas de sobrevivencia son distintas dada la condición laboral. A continuación se presenta la información que sustenta la anterior afirmación:



Gráfica 5. Función de sobrevivencia por trabajaba en DI

La probabilidad de permanecer en Diseño Industrial es superior en quienes no trabajan que en los que si lo hacen como es evidente en la gráfica 5, por ejemplo en el semestre uno, para el caso de los estudiantes que no trabajaban la probabilidad es de 88.9% y para aquellos estudiantes que si trabajaron al presentar el examen de estado la probabilidad es cercana al 83%.

Sin embargo, los resultados de esta variable y su consideración en el modelo de Cox para deserción (se presenta más adelante) se deben tomar con precaución pues en primer lugar, no implica necesariamente el estado actual del estudiante ya que su recolección corresponde a un periodo anterior al ingreso a la universidad, lo que se puede deducir es el hecho que para el caso de un estudiante que trabaja su responsabilidad en los estudios es mayor.

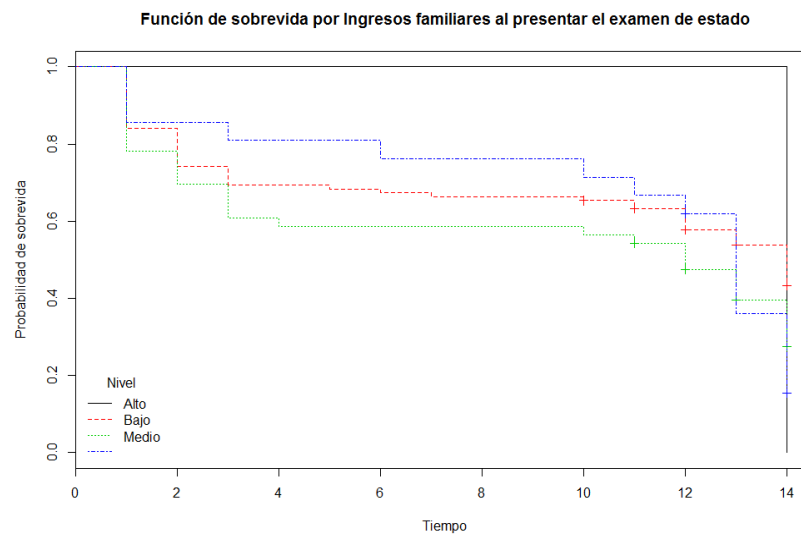
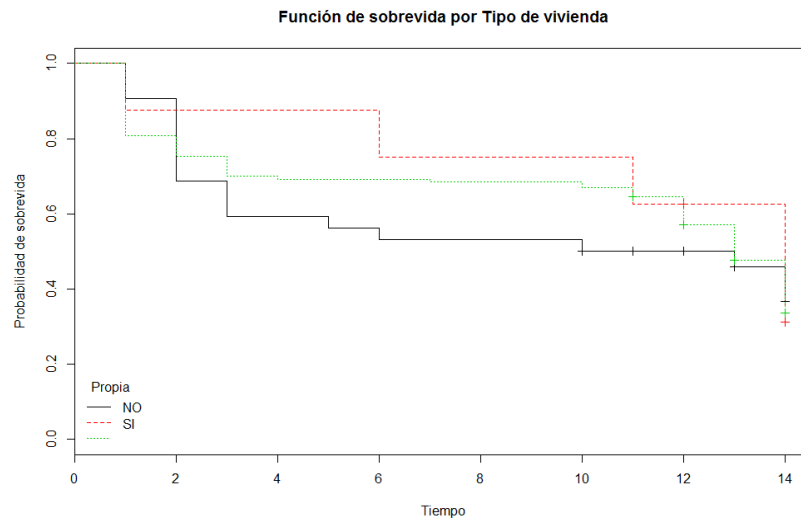
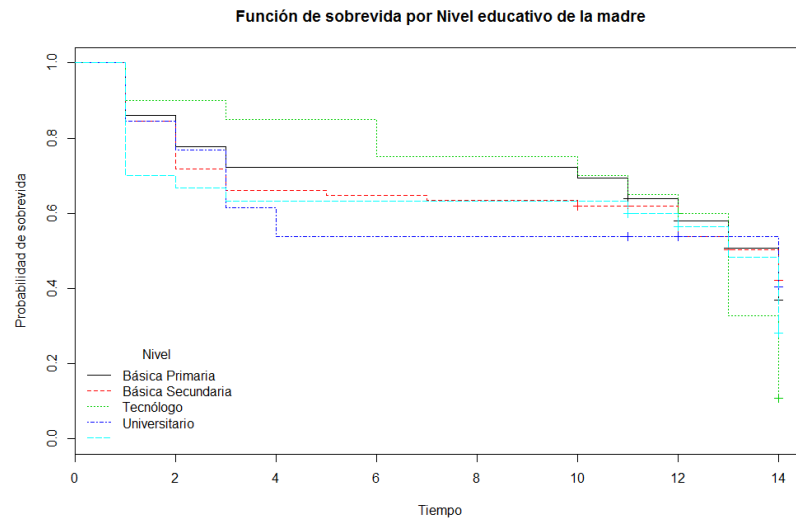
Además, de los 170 estudiantes del estudio tan sólo seis presentaron la condición de estar trabajando al momento de presentar el examen de estado (ICFES o prueba SABER 11).

Tabla 11. Resumen estimador Kaplan Meier por trabajaba en DI

Trabajaba=NO					95% IC	
Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	146	26	0.822	0.0317	0.762	0.886
2	120	11	0.747	0.0360	0.679	0.821
3	109	8	0.692	0.0382	0.621	0.771
4	101	1	0.685	0.0384	0.614	0.765
5	100	1	0.678	0.0387	0.606	0.758
6	99	1	0.671	0.0389	0.599	0.752
7	98	1	0.664	0.0391	0.592	0.746
10	97	2	0.651	0.0395	0.578	0.733
11	94	2	0.637	0.0398	0.563	0.720
12	84	8	0.576	0.0414	0.501	0.663
13	66	7	0.515	0.0430	0.437	0.607
14	52	13	0.386	0.0447	0.308	0.485
Trabajaba=NA					95% IC	
Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	18	2	0.889	0.0741	0.7549	1.000
3	16	1	0.833	0.0878	0.6778	1.000
6	15	1	0.778	0.0980	0.6076	0.996
10	14	1	0.722	0.1056	0.5423	0.962
11	13	1	0.667	0.1111	0.4809	0.924
12	12	1	0.611	0.1149	0.4227	0.883
13	10	4	0.367	0.1171	0.1961	0.686
14	6	4	0.122	0.0806	0.0335	0.445
Trabajaba=SI					95% IC	
Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	6	1	0.833	0.152	0.5827	1.000
2	5	3	0.333	0.192	0.1075	1.000
3	2	1	0.167	0.152	0.0278	0.997
11	1	1	0.000	NaN	NA	NA

Fuente: La Autora

En el Anexo C se presentan las sentencias para las pruebas de hipótesis para nivel educativo de la madre, ingreso familiar y tenencia de vivienda, se encontró que para las diferentes categorías de las variables la sobrevivencia (a la deserción) tiene el mismo comportamiento. Las funciones de supervivencia se presentan en seguida:



Gráfica 6. Otras curvas de sobrevivencia de DI

#### 4.2.2. Modelo de COX – Factores que influyen en la deserción

Para medir los efectos de las variables consideradas en el estudio como explicativas del fenómeno de la deserción (género, edad de ingreso al programa, nivel educativo de la madre al momento de presentar el examen de estado, número de hermanos, vivienda propia, trabajaba al momento de presentar el examen de estado, ingreso familiar al presentar el examen de estado, puntaje ICFES estandarizado), en seguida se presentan las estimaciones del modelo de riesgo proporcional utilizando el modelo semiparamétrico de Cox. Los resultados del modelo inicial, es decir con todas las variables, se presentan en la tabla 12.

Tabla 12. Modelo de Cox para la deserción en DI

	Coeficiente regresión	Exp (Coeficiente regresión)	Error estándar (Coeficiente regresión)	z	Pr(> z )
Edad.de.ingreso.al.programa	-0.036730	0.963937	0.048566	-0.756	0.4495
Genero[T.M]	-0.153232	0.857930	0.307827	-0.498	0.6186
Ingreso.Familiar [T.BJ]	-0.360012	0.697668	0.859632	-0.419	0.6754
Ingreso.Familiar [T.ME]	-0.175508	0.839031	0.786273	-0.223	0.8234
Nivel.Educativo.Madre[T.BS]	-0.041127	0.959707	0.346983	-0.119	0.9056
Nivel.Educativo.Madre[T.TEC]	-0.378091	0.685168	0.562343	-0.672	0.5014
Nivel.Educativo.Madre[T.UN]	0.128131	1.136.701	0.525847	0.244	0.8075
Numero.de.Hermanos	0.059986	1.061.822	0.147182	0.408	0.6836
Puntaje	-0.004170	0.995839	0.006418	-0.650	0.5158
Trabajaba [T.SI]	1.362.517	3.906.014	0.616362	2.211	0.0271 *
Vivienda.Propia[T.SI]	-0.256986	0.773379	0.319532	-0.804	0.4212

Fuente: La Autora

En seguida procede a seleccionar el modelo más parsimonioso, para esto se utiliza el método hacia adelante con el criterio AIC (Criterio de Información de Akaike), encontrándose que el modelo óptimo queda determinado por trabajaba al momento de presentar el examen de estado, se decide incluir la variable edad de ingreso al programa teniendo en cuenta la revisión de literatura pues en estudios realizados por Giovanoly (2002) y Barrera (2009) esta variable es relevante a la hora de

explicar el tiempo hasta que un estudiante deserta. La tabla 13 resume el modelo óptimo.

Tabla 13. Resumen modelo óptimo para la deserción en DI

	Coeficiente regresión	Exp(Coeficiente regresión)	Error estándar (Coeficiente regresión)	z	Pr(> z )
EdadIngreso	-0.02635	0.97399	0.04427	-0.595	0.55172
Trabaja [T.SI]	1.46920	4.34577	0.48144	3.052	0.00228 **
Signif. codes:	0 ****	0.001 ***	0.01 **	0.05 *	0.1
		Exp(Coeficiente regresión)	Exp(-Coeficiente de regresión)	L. inferior 0.95	L.superi or0.95
EdadIngreso		0.974	1.0267	0.893	1.062
Trabaja [T.SI]		4.346	0.2301	1.691	11.165

Fuente: La Autora

De la tabla 13 se deduce:

Edad de ingreso al programa: La estimación del riesgo relativo respecto a la edad de ingreso al programa es 0.97399, la estimación puntual es inferior a 1, indica que por cada año más en la edad de ingreso al programa disminuye el riesgo de desertar, concretamente, dejando la otra variable constante, si se comparan dos estudiantes, para aquel estudiante menor en un año, se tiene 1.0267 veces más probabilidad de desertar que el estudiante mayor. Observamos que el p-valor es 0.55172 (mayor que 0.05), entonces se tiene que la relación entre la edad de ingreso al programa y el riesgo de desertar no son estadísticamente significativa.

Trabajaba al presentar el examen de Estado: El riesgo relativo de trabajar al presentar el examen de estado respecto al no trabajar es de 4.34577. Con lo cual tiene más riesgo de desertar los que trabajan que aquellos que no lo hacen, es decir cuando un estudiante no trabaja se multiplica por 0,2301 la probabilidad de desertar, es decir disminuye respecto al que trabaja. Esta relación si es significativa (el p-valor es 0.00228, menor que 0.05).

De acuerdo con lo anterior se concluye que trabajar al momento de presentar el examen de Estado afecta negativamente en el tiempo de sobrevivencia, por ello se incrementa el riesgo de desertar.

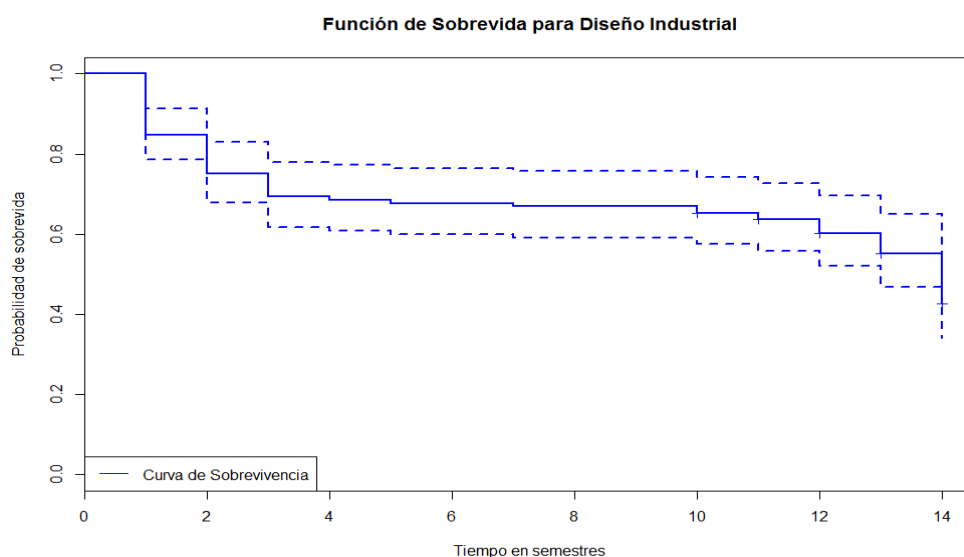
El modelo es aceptable para cualquiera de los tres criterios que se usan y las variables seleccionadas en el modelo óptimo explican el 5.3% de la variabilidad que se presenta en el tiempo del fenómeno de la deserción.

Tabla 14. Estimación de los parámetros del modelo de Cox de deserción en DI

Concordancia= 0.554 (se=0.042)			
Rsquare= 0.053 (max possible= 0.99 )			
Likelihood ratio test=	6.7	con 2 df,	p=0.03501
Wald test =	9.4	con 2 df,	p=0.009114
Score (logrank) test =	11.04	con 2 df,	p=0.004008

Fuente: La Autora

Se hace necesario determinar la función de sobrevivencia para el modelo ajustado. Tal estimación se presenta a continuación por medio de la gráfica y la tabla con los datos de la estimación:



Gráfica 7. Función de sobrevivencia para el modelo de deserción en DI

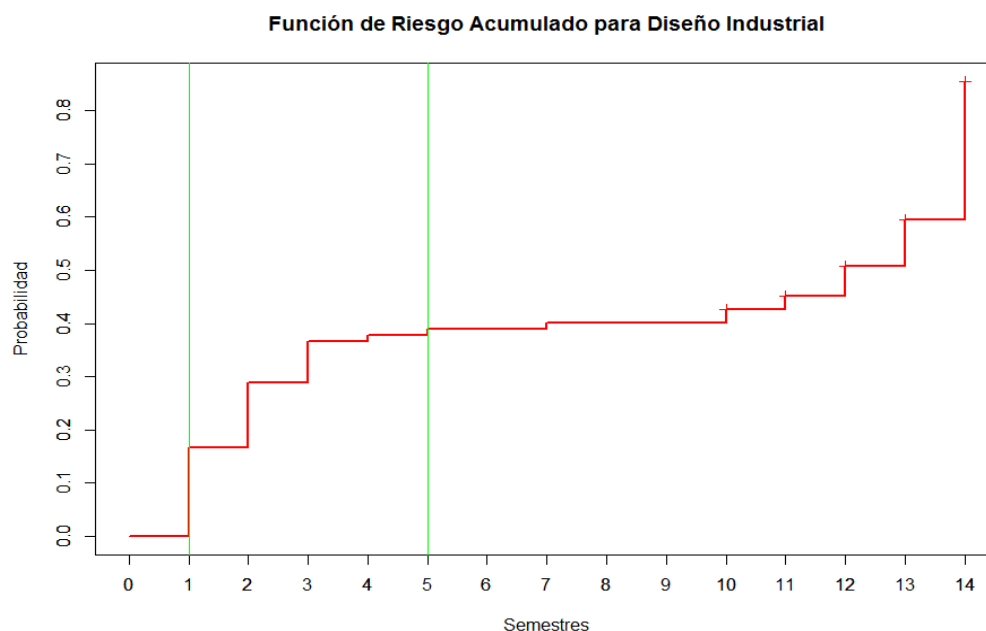
Analizando la primera línea de la tabla 15 con las estimaciones del modelo ajustado, se encuentra que la probabilidad de desertar fue mayor en la deserción precoz y temprana, en el primer semestre la probabilidad de no desertar considerando las variables explicativas es 84.7%, entre el cuarto y séptimo semestre se genera una estabilidad en la probabilidad alrededor del 67%, aunque se evidencia un crecimiento después del décimo semestre con lo cual la probabilidad de sobrevivir es cercana al 42.5%. En los semestres seis, ocho y nueve no se presentó el evento de deserción.

Tabla 15. Estimaciones función de sobrevivencia del modelo de deserción en DI

Semestre	Número de individuos	Número de fallas	Sobrevivencia	Error estándar	Límite inferior	Límite superior
1	124	19	0.847	0.0323	0.786	0.913
2	105	12	0.750	0.0389	0.678	0.830
3	93	7	0.694	0.0414	0.617	0.780
4	86	1	0.685	0.0417	0.608	0.772
5	85	1	0.677	0.0420	0.600	0.765
7	84	1	0.669	0.0422	0.591	0.757
10	83	2	0.653	0.0427	0.575	0.743
11	80	2	0.637	0.0432	0.558	0.727
12	73	4	0.602	0.0442	0.521	0.695
13	60	5	0.552	0.0459	0.469	0.649
14	48	11	0.425	0.0487	0.340	0.532

Fuente: La Autora

A partir de la gráfica del riesgo acumulado, se puede afirmar que la probabilidad de desertar justo al terminar el primer semestre es del 16%, la probabilidad de desertar justo al terminar el segundo semestre es del 12.76% y en el caso del tercer semestre la probabilidad es del 7.76%, para los siguientes semestres el riesgo instantáneo es cercano al 1% hasta el semestre número doce, para el semestre número trece la probabilidad de desertar justo al terminarlo es del 5%.



Gráfica 8. Función de riesgo acumulado del modelo de Cox de deserción en DI

### 4.2.3. Evaluación del modelo de COX

Con el fin de comprobar si el modelo ajustado para la deserción en el programa de Diseño Industrial es adecuado, se prueba el supuesto de riesgos proporcionales del modelo de regresión a través de R-Commander, obteniendo la siguiente tabla:

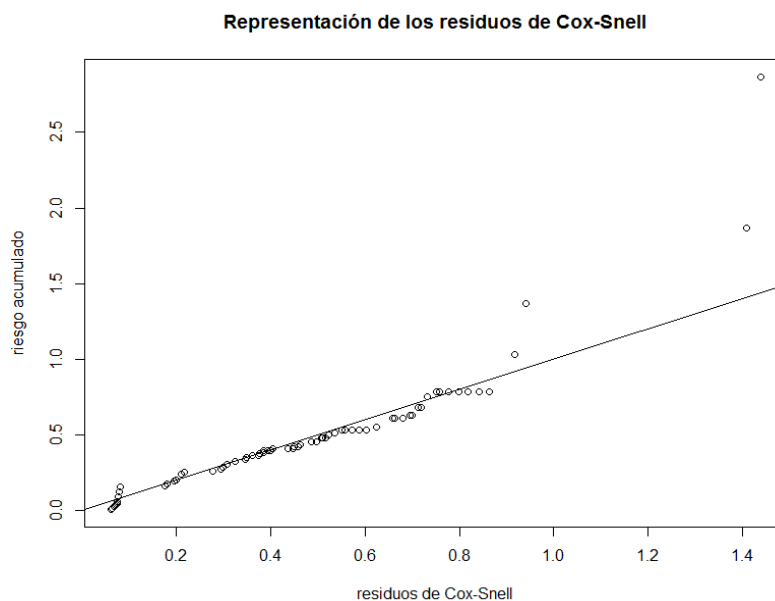
Tabla 16. Riesgos proporcionales de deserción en DI

	rho	Chisq	p
Trabajaba[T.SI]	0.1696	2.008	0.156
Edad.de.ingreso.al.programa	-0.0828	0.456	0.500
GLOBAL	NA	2.294	0.318

Fuente: La Autora

De acuerdo con la anterior tabla se afirma que no existen evidencias significativas al 5% de que se viole el supuesto de riesgos proporcionales para ninguna de las dos covariables ni globalmente, pues, los p-valores son mayores al 0.05.

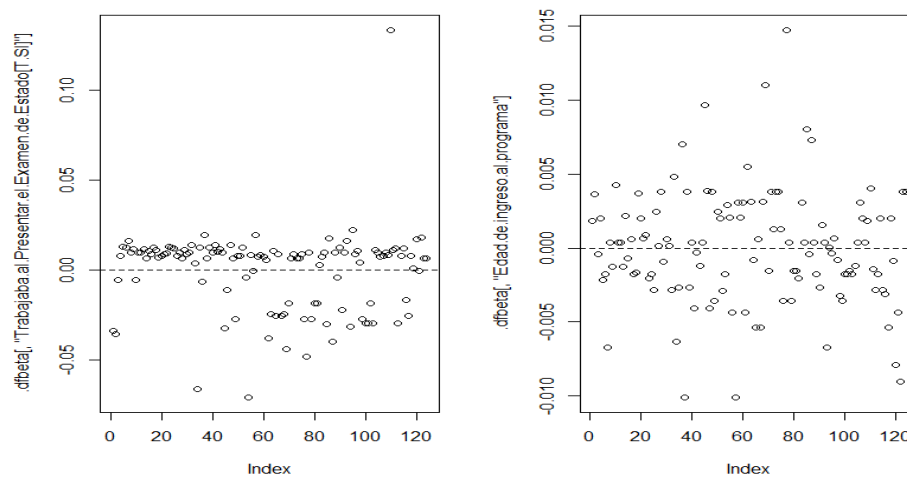
Se realiza la comprobación de la hipótesis global calculando los residuos de Cox-Snell, es decir si el modelo es correcto y la estimación de los  $\beta$  son cercanas a los valores reales, como el estimador debería seguir una recta con pendiente igual a la unidad, como se observa en la gráfica 9, se determina que el modelo ajusta bien los datos.



Gráfica 9. Residuos de Cox-Snell para deserción en DI



En el proceso de comprobación de la hipótesis sobre cada observación en el modelo, se usan los residuos dfbeta, específicamente se analiza la influencia de cada observación en el modelo, para identificar lo anterior se presenta la gráfica 10, con la representación de la observación por el cambio de cada escala aproximada del coeficiente, (dividiendo por el error estándar), después de eliminar la observación del modelo.



Gráfica 10. Residuos escalonados de Schoenfeld de deserción en DI

Los residuos de la anterior figura están centrados cerca al origen para el caso de ocupación laboral (trabajar), y no presentan patrones definidos. Se nos presentan datos demasiados alejados del origen en la covariable edad de ingreso a este programa.

#### 4.2.4. Pronóstico a partir del modelo

La estimación o ajuste del modelo de riesgos proporcionales está dado por la expresión  $h_i(t, X, \beta) = h_0(t) \exp(X^T \beta)$ . Las dos componentes del modelo  $h_0(t)$  (función de riesgo inicial) y  $\exp(X^T \beta)$  pueden ser estimadas por separado, se estiman primero los componentes del vector de parámetros y a partir de ellos se construye el estimador de la función de riesgo inicial. Nótese que la función de riesgo inicial depende del tiempo mientras que la segunda componente depende únicamente de las covariables. Aunque, en el modelo de riesgos proporcionales de Cox se debe tener en cuenta que los riesgos para dos conjuntos diferentes de valores de las covariables conservan la misma proporción a lo largo del tiempo, de ahí su denominación.

Con base en los resultados obtenidos hasta el momento, se tiene que el modelo estimado corresponde a:

$$h_i(t, X, \beta) = h_0(t) \exp(-0.02635(Edadingreso)) + 1.4692(Trabajar[T.SI])$$

El exponente del modelo de Cox, que particularmente para el estudio corresponde a la anterior ecuación, se denomina *índice pronóstico* ( $P_I$ ), (referencia). De manera que la tasa de riesgo se puede expresar a través de este indicador en lugar de utilizar los valores del estudiante para cada variable. En la práctica es mejor usar el *índice de pronóstico centrado* ( $P_{IC}$ ), que se obtiene centrando las variables predictoras:

$$P_{IC} = \beta_1(X_1 - \bar{X}_1) + \dots + \beta_p(X_p - \bar{X}_p)$$

Que para nuestro caso corresponde a:

$$P_{IC} = -0.02635(Edadingreso - 19.09) + 0(Trabajar[T.NO])$$

Un índice pronóstico centrado es igual a cero cuando corresponde en el caso del programa de Diseño Industrial a un estudiante que no trabajaba al momento de presentar su examen de estado y cuya edad promedio es de 19 años (estudiante base).

La diferencia entre los índices pronósticos de dos estudiantes permite estimar su riesgo relativo (RR), veamos:

$$\begin{aligned} RR &= \frac{h_i(t, X_i, \beta)}{h_j(t, X_j, \beta)} = \frac{h_0(t) \exp(P_{IC\text{individuo}i})}{h_0(t) \exp(P_{IC\text{individuo}j})} \\ &= \exp(P_{IC\text{individuo}i} - P_{IC\text{individuo}j}) \end{aligned}$$

Teniendo en cuenta esto a continuación se presenta el RR, respecto al estudiante base, para cinco estudiantes, seleccionados aleatoriamente, y que cursaron el primer semestre del programa, es decir, estudiantes que ingresaron en el primer semestre de 2015:

Tabla 17. Riesgo relativo de deserción en DI

Estudiante	Variables		$P_{IC}$	RR respecto a un estudiante promedio
	Edad ingreso	Trabajar		
1	17	Si	1.5242	4.5914
2	17	No	0.055	1.73
3	19	No	0.0023715	1.0023
4	18	No	0.0287215	1.0291
5	17	No	0.055	1.73

Fuente: La Autora

Lo anterior significa que el riesgo de deserción para los estudiantes con  $P_{IC} = 0.055$ , es 1.73 superior que para el estudiante base del Programa, es decir un estudiante que no trabajaba al momento de presentar su examen de estado y cuya edad promedio es de 19 años.

El riesgo de deserción para los estudiantes con  $P_{IC} = 1.5242$ , es 4.5914 superior que para el estudiante base del Programa.

En el caso del riesgo de deserción para los estudiantes con  $P_{IC} = 0.0023715$ , es 1.0023 superior que para el estudiante base del Programa, es decir tiene riesgo similar de desertar.

En el caso del riesgo de deserción para los estudiantes con  $P_{IC} = 0.0287215$ , es 1.0291 superior que para el estudiante base del Programa.

A continuación se presenta el mismo proceso de acuerdo al modelo de graduación para este programa.

### 4.3. MODELO PARA GRADUACIÓN

Como la deserción se encuentra ligada al hecho de graduarse o no, a continuación se analiza si las variables usadas como explicativas en el modelo de deserción explican la graduación en el programa académico Diseño Industrial.

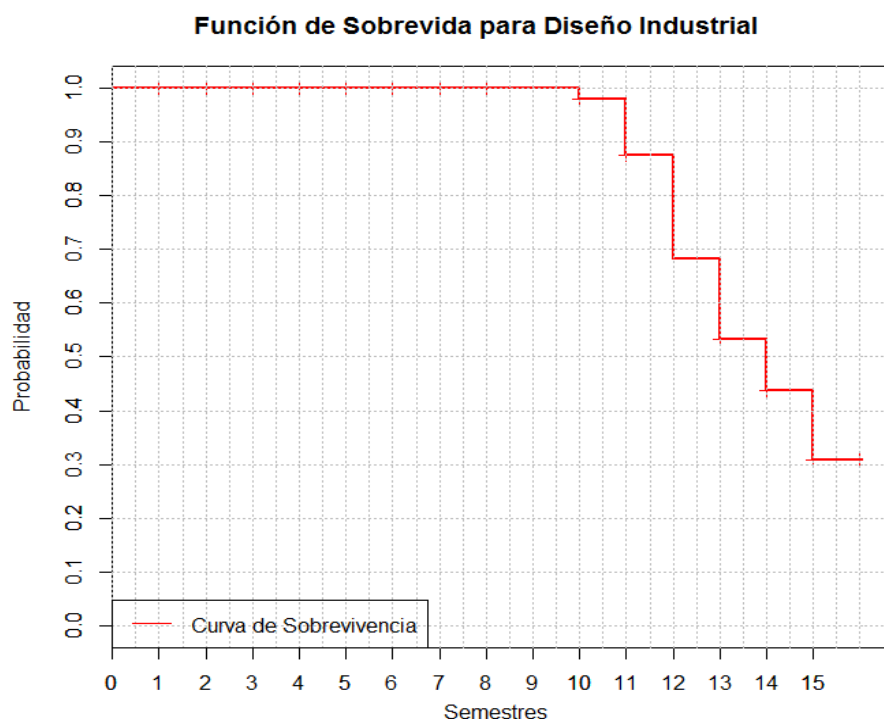
Los cambios que se generan en la caracterización de los datos de sobrevivencia son:

- Evento de interés o falla: Terminación del programa Diseño Industrial por parte del estudiante (graduación).

- La variable respuesta: Tiempo hasta que el estudiante finaliza este programa, lo cual se mide en número de semestres cursados hasta que se gradúa.
- Tipo de censura: Tipo I y a derecha, los individuos ingresan al estudio en diferentes tiempos, como ya se había mencionado, diferentes cohortes, y el punto final del estudio es el mismo para todos. Se considera como censura al estudiante que deserta (causas académicas o no académicas) o que continúa estudiando.
- Tiempo de censura: Tiempo hasta el percentil 75 que corresponde al semestre número quince.

#### 4.3.1. Función de sobrevivencia

Se usa a R-Commander para realizar la estimación de la función de Sobrevivencia por medio del método de Kaplan Meier, a continuación se presenta la gráfica obtenida:



Gráfica 11. Función de sobrevivencia estimada para graduación en DI

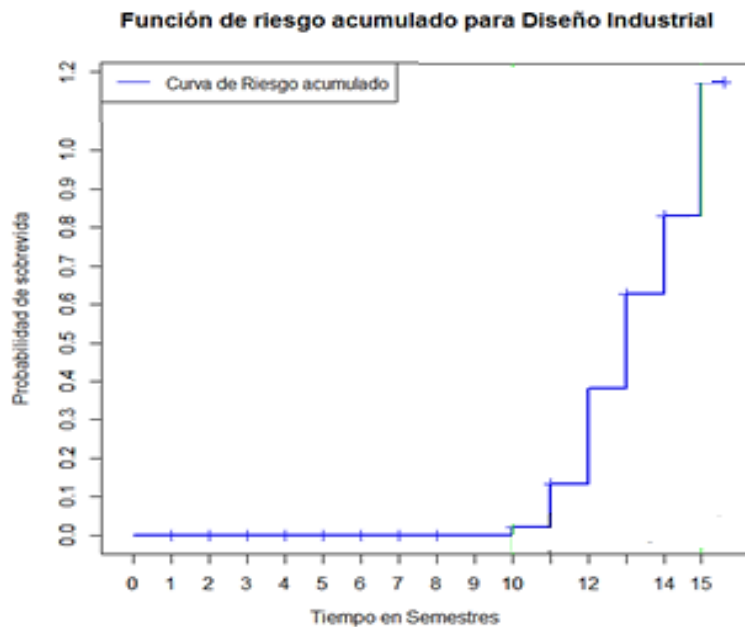
La tabla 18 presenta las estimaciones de la función de sobrevivencia para el evento de graduación sin tener en cuenta ninguna variable, los eventos para este caso se encuentran desde el décimo semestre, el semestre número doce presenta la mayor cantidad de eventos y además se observa que la probabilidad de no graduarse es del 30.8%.

Tabla 18. Estimaciones función de sobrevivencia para graduación en DI

Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
10	49	1	0.980	0.0202	0.941	1.000
11	47	5	0.875	0.0476	0.787	0.974
12	41	9	0.683	0.0677	0.563	0.830
13	32	7	0.534	0.0727	0.409	0.697
14	22	4	0.437	0.0739	0.313	0.609
15	17	5	0.308	0.0711	0.196	0.484

Fuente: La Autora

A partir de la función de riesgo acumulada se puede afirmar que la probabilidad de que un estudiante del programa se gradúe justo al terminar el décimo semestre es del 2%, en el semestre once la probabilidad es del 11.35%, se evidencia que en el semestre, justo al terminar el semestre número catorce la probabilidad de graduarse es del 20.05%, de acuerdo a ello se determina que a medida que aumentan los semestres aumenta la probabilidad de la terminación del programa académico o graduarse.



Gráfica 12. Función de riesgo acumulado para graduación de DI

#### 4.3.2. Modelo de COX – Factores que influyen en la graduación

La función de sobrevivencia se describe a continuación, en primer lugar se calculan las estimaciones del modelo semiparamétrico de Cox. Los resultados del modelo con todas las variables se presentan en la tabla 19.

Se determina el modelo más parsimonioso, usando el método hacia adelante y teniendo como criterio el AIC (Criterio de Información de Akaike), encontrándose que la variable materias repetidas explica la graduación en el programa. En el resumen se observa la estimación del riesgo relativo para la covariable, su intervalo de confianza y el p-valor sobre su significancia, además se tiene la inversa del riesgo relativo para los casos inferiores a 1.

A continuación se presenta un resumen del modelo óptimo en la tabla 20, el cual sólo incluye la variable número de materias perdidas.

Tabla 19. Modelo de Cox para graduación en DI

	Coeficiente regresión	Exp (Coeficiente regresión)	Error estándar (Coeficiente regresión)	z	Pr(> z )
Edad.de.ingreso.al.programa	-0.1077931	0.8978133	0.0824015	-1.308	0.1908
Genero[T.M]	-0.7858248	0.4557436	0.4165821	-1.886	0.0592
Ingreso.Familiar[T.BJ]	-0.5672138	0.5671033	0.7179257	-0.790	0.4295
Ingreso.Familiar[T.ME]	NA	NA	0.0000000	NA	NA
Materias.Perdidas	-0.0725025	0.9300634	0.1139571	-0.636	0.5246
Materias.Repetidas	-0.0410721	0.9597599	0.1217648	-0.337	0.7359
Nivel.Educativo.Madre[T.BS]	-0.1571648	0.8545632	0.5159410	-0.305	0.7607
Nivel.Educativo.Madre[T.TEC]	-0.1304294	0.8777185	0.6919237	-0.189	0.8505
Nivel.Educativo.Madre[T.UN]	-1,62E+07	0.1981393	0.8961430	-1.806	0.0709
Numero.de.Hermanos	0.0015175	10.015.186	0.2485778	0.006	0.9951
Puntaje	-0.0009816	0.9990189	0.0134463	-0.073	0.9418
Trabajaba[T.SI]	NA	NA	0.0000000	NA	NA
Vivienda.Propia[T.SI]	-0.5315900	0.5876698	0.5407316	-0.983	0.3256

Fuente: La Autora

De acuerdo a la siguiente tabla, se puede deducir que para la variable número de materias perdidas se tiene una estimación del riesgo relativo del 0.89013, la estimación puntual es inferior a 1, lo cual indica que si se comparan dos estudiantes, aquel que pierda una materia menos que otro tiene 1.123 veces más probabilidad

de graduarse. Otro dato importante que se observa es el p-valor 0.00145 (menor que 0.05), con ello se identifica que la relación entre el número de las materias perdidas y el tiempo en graduarse es estadísticamente significativa.

Tabla 20. Resumen modelo óptimo para graduación en DI

	Coeficiente regresión	Exp(Coeficiente regresión)	Error estándar (Coeficiente regresión)	z	Pr(> z )	
Materias.Perdidas	-0.11639	0.89013	0.03655	- 3.184,00	0.00145	**
--- Signif. codes: 0 '***' 0.001 '***' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
	Exp(Coeficiente regresión)	Exp(-Coeficiente de regresión)	L. inferior 0.95	L.superior0.95		
Materias.Perdidas	0.89013	1.123	0.8286	0.9562		

Fuente: La Autora

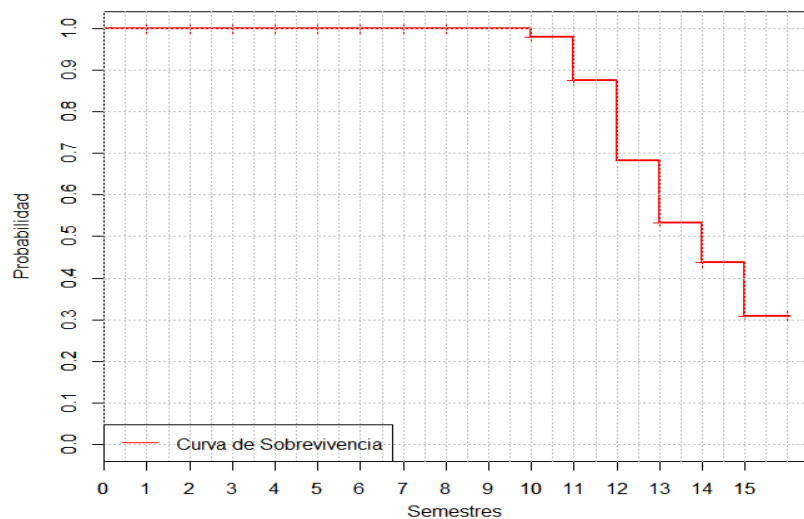
El modelo de graduación es aceptable con cualquiera de los tres criterios que se usan, debido a que se aprecia un p-valor significativamente pequeño, lo cual evidencia que la variable materias perdidas es una variable explicativa con un coeficiente significativamente distinto de cero, esta variable logra explicar el 10% de la variabilidad que se presenta en el tiempo hasta graduarse, lo anterior se afirma según la tabla 21.

Tabla 21. Estimación del modelo de Cox de graduación en DI

Concordancia= 0.697 (se=0.067)			
Rsquare= 0.109 (max possible= 0.861 )			
Likelihood ratio test=	12.27	con 1 df,	p=0.0004592
Wald test =	10.14	con 1 df,	p=0.001452
Score (logrank) test =	11.06	con 1 df,	p=0.0008839

Fuente: La Autora

Continuando con el análisis del modelo ajustado para la graduación en el programa de estudio, se determina la función de sobrevida, esta estimación se presenta a continuación por medio de una gráfica y la tabla, lo cual se realizó por el método de Kaplan-Meier.



Gráfica 13. Función de sobrevivencia para el modelo de graduación en DI

A partir de la gráfica 13 y de acuerdo con la tabla 22, se deduce que la probabilidad de no graduarse en el semestre diez es del 98%, empieza a disminuir la probabilidad con lo cual hasta el semestre doce la probabilidad de no graduarse a disminuido en 29.7% es decir que en ese semestre es del 68.3%, en el caso del semestre catorce la probabilidad de no graduarse es cercana al 43% y finalmente en el semestre quince se puede concluir que es más probable que suceda la graduación pues la probabilidad de no graduarse es cercana al 30%, es decir se puede afirmar que la mayoría de estudiantes que llegan al semestre quince logran graduarse del programa académico de DI.

Tabla 22. Estimaciones función de sobrevivencia del Modelo de graduación en DI

Semestre	Número de individuos	Número de fallas	Sobrevencia	Error estándar	Límite inferior	Límite superior
10	49	1	0.980	0.0202	0.941	1.000
11	47	5	0.875	0.0476	0.787	0.974
12	41	9	0.683	0.0677	0.563	0.830
13	32	7	0.534	0.0727	0.409	0.697
14	22	4	0.437	0.0739	0.313	0.609
15	17	5	0.308	0.0711	0.196	0.484

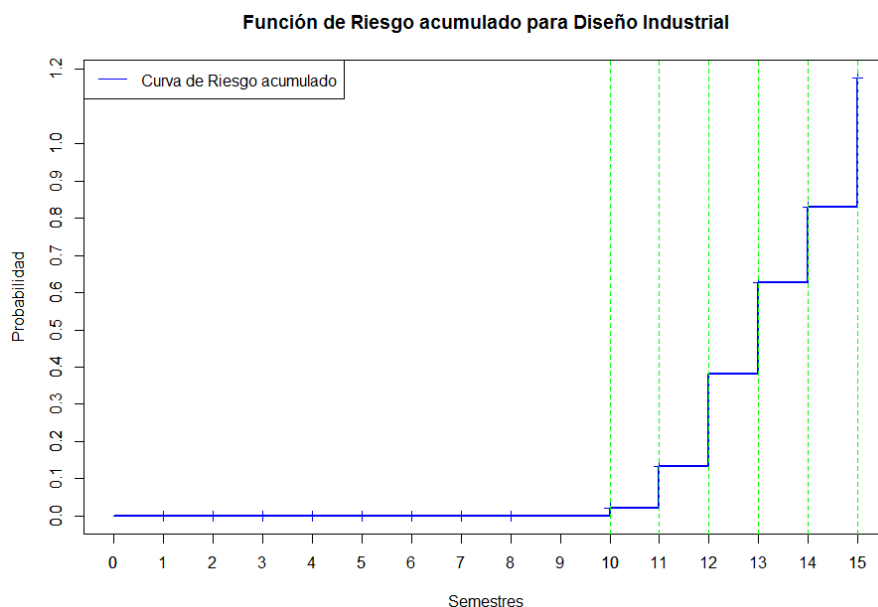
Fuente: La Autora

En seguida se tiene la función de riesgo acumulado, a partir de la cual se infiere que la probabilidad de graduarse justo al terminar el décimo semestre es del 2%, en el décimo primer semestre la probabilidad de graduarse al terminar ese semestre es del 11.35%, al observar el semestre doce se obtuvo una probabilidad 24.77%, en el



semestre número catorce la probabilidad de graduarse justo al terminar ese semestre es del 20.05%.

Con lo anterior se ratifica lo mencionado en la curva de sobrevivencia, el aumento de la probabilidad de graduarse a medida que incrementan los semestres de estudio, lo cual es evidente en la gráfica 14, pues se observan los incrementos de probabilidad de graduarse de un semestre a otro.



Gráfica 14. Función riesgo acumulado del modelo de Cox de graduación en DI

### 4.2.3. Evaluación del modelo de COX

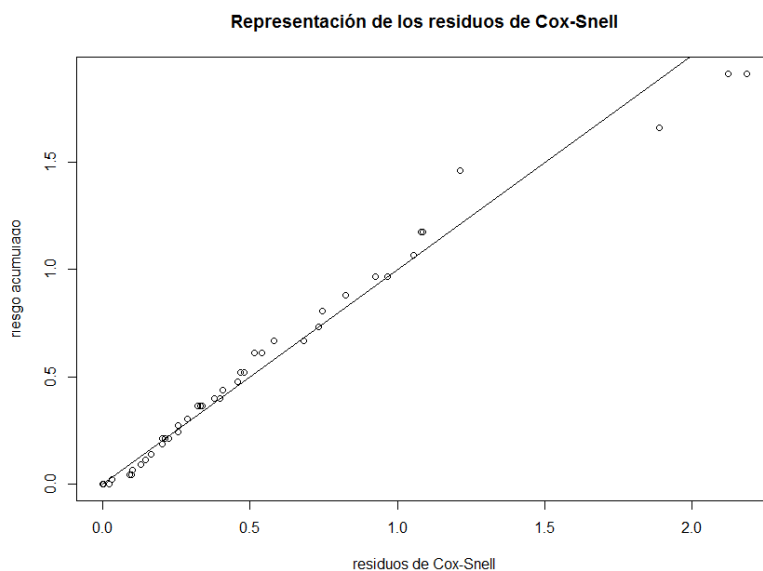
Esta sección tiene como objetivo comprobar si el modelo ajustado en el caso de la graduación en el programa Diseño Industrial es adecuado, a continuación se prueba el supuesto de riesgos proporcionales del modelo de regresión a través de R-Commander, obteniendo la tabla 23, a partir de la cual se determina la no existencia de evidencias significativas al 5% de que se viole el supuesto de riesgos proporcionales para la covariable, ya que el p-valor es mayor al 0.05.

Tabla 23. Riesgos proporcionales de graduación en DI

	rho	chisq	p
Materias.Perdidas	0.222	1.04	0.308

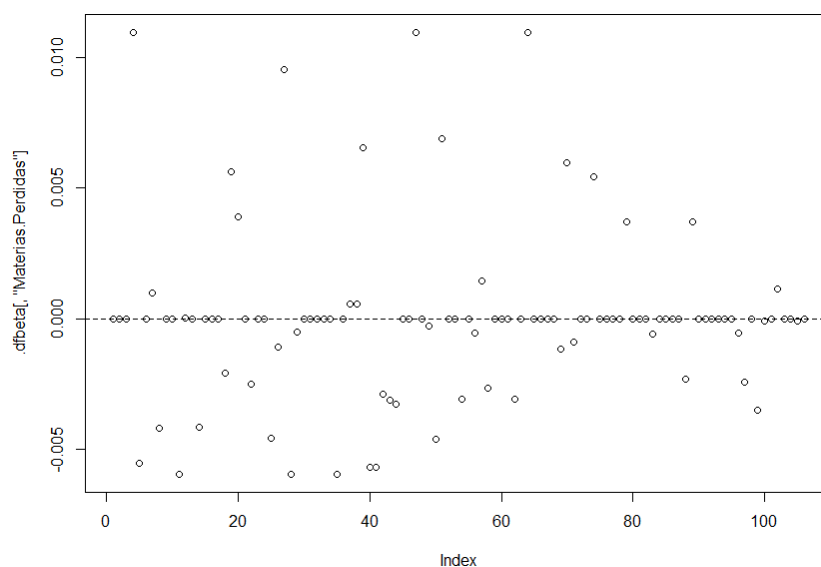
Fuente: La Autora

Se continúa con la comprobación de la hipótesis global por medio de los residuos de Cox-Snell, entonces se verifica si la estimación de los  $\beta$  es cercana a los valores reales, como el estimador debería seguir una recta con pendiente igual a la unidad, se observa la siguiente gráfica, se determina que el modelo ajusta bien los datos.



Gráfica 15. Residuos de Cox-Snell para graduación en DI

En la comprobación de la hipótesis sobre la observación en el modelo de graduación de acuerdo a este programa, en esta evaluación se usan los residuos  $dfbeta$ , es decir se analiza la influencia de cada observación en el modelo, lo anterior se observa en la siguiente ilustración (gráfica 16).



Gráfica 16. Residuos escalonados de Schoenfeld de graduación en DI

Los residuos escalonados de Schoenfeld ubicados en la anterior gráfica están centrados cerca al origen para el caso de materias perdidas, en otras palabras no se presentan problemas respecto a esta variable.

Como conclusión principal del programa de Diseño Industrial se tiene que el 60% de los estudiantes que ingresan desertan, cifra que según el MEN (2014) está por encima del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2% e incluso la del área básica de conocimiento del Programa – Ingeniería, arquitectura, urbanismo y Afines - que se encuentra en 50,7%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 60, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

La mayoría deserta por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Desertan en su mayoría los hombres, características sobresalientes de los estudiantes que ingresaron a Diseño Industrial son la edad promedio de ingreso (17 años), el 65% de los estudiantes contaban con vivienda propia, el 84% no trabajaban al momento de presentar el examen de estado, el puntaje promedio al momento de presentar el examen de estado fue de 67.6, el 35,6% de los estudiantes tenían madres con nivel educativo de básica primaria y el 44,5% de los estudiantes tenían dos hermanos al momento de presentar el examen de Estado.

Para el programa de DI, la posibilidad de que un estudiante se vaya en primer semestre es del 19%, donde el periodo crítico con mayor intensidad es de manera temprana (del primero al tercer semestre). Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas.

El factor que aumenta el riesgo de desertar en DI es trabajar al momento de presentar su examen de estado, lo cual se puede atribuir a la falta de tiempo con las actividades académicas, es decir que en la medida que un estudiante no trabaje se reduce el riesgo de deserción, se recomienda a la Universidad y al programa crear políticas de patrocinio y apoyo a aquellos estudiantes con esta característica. Se disminuye el riesgo de desertar al presentar mayor edad al ingresar al programa,

lo cual puede evidenciarse debido al tema vocacional en los estudiantes que ingresan muy jóvenes a DI.

Por el contrario, para la graduación el factor que disminuye esta propensión es mayor número de asignaturas perdidas, luego un estudiante presenta menor probabilidad de graduarse. Además a partir del semestre número catorce es más probable que un estudiante se gradúe en el 14, hecho que es preocupante ya que gasta cuatro semestres más de lo previsto para obtener el título profesional y posiblemente entre más tiempo transcurra será difícil culminar con el proyecto educativo. Por lo anterior se sugirieren actividades para aumentar los niveles de graduación inmediatamente se culmine con la respectiva carga académica, por ejemplo la asignación de temas de grado en semestres inferiores, vinculación inmediata con algún programa de posgrado según la normatividad de la Universidad, o en la pasantía involúcralos con proyectos productivos.

## 5. MODELO SOBREVIDA PARA INGENIERIA ELECTROMECAÁNICA

En este capítulo se tiene en primer lugar la descripción de las variables de acuerdo a la muestra seleccionada, luego de las variables escogidas como explicativas por medio de un análisis univariado y bivariado, en segundo lugar la descripción del proceso de la construcción del modelo de sobrevida para Ingeniería Industrial de acuerdo a la deserción y a la graduación.

### 5.1. DESCRIPCIÓN DE LOS DATOS

En esta sección se describen las variables estudiadas, resumidas en porcentajes y con algunos estadísticos como la media, la mediana, mínimo y máximo del grupo poblacional. También, se presenta un análisis bivariado desde la intención del modelamiento, tanto la descripción como los análisis se realizan a través de R-Commander. Todos los detalles de las sentencias en R figuran en el Anexo C. En la tabla 24 figuran las estadísticas de las variables:

Tabla 24. Resumen descriptivo de las variables de estudio de ING

<b>Variable</b>	<b>Unidades de medición /codificación</b>	<b>Resumen numérico</b>
<i>Edad de Presentación del Examen de Estado</i>	En Años	Media = 16.95 Min – Max = 12 – 26 Mediana = 16
<i>Estado del estudiante</i>	Matriculado= MT Desertor= DE Graduado= GR Terminación Académica= TA  Sin Clasificar= SC	19 (10.61%) 78 (43.57%) 71 (40.22%) 11 (6.14%)  2 (3.35%)
<i>Ingreso Familiar al Presentar el Examen de Estado</i>	Bajo= [0-1) y [1-2) SMMLV Medio= [2-3) y [3-5) SMMLV Alto= [5-7) y [10-) SMMLV  Sin Clasificar= SC	126 (70.39%) 23 (12.84%) 9 (5.02%)  21 (11.73%)
<i>Género</i>	F= Femenino M= Masculino	17 (9.5%) 162 (90.5%)

<b>Variable</b>	<b>Unidades de medición /codificación</b>	<b>Resumen numérico</b>
<i>Nivel educativo de la madre</i>	Básica Primaria= BP Básica Secundaria= BS Técnico= TEC Universitario= UN Sin Clasificar= SC	56 (31.28%) 70 (39.1%) 18 (19.05%) 14 (7.82%) 21 (11.73%)
<i>Número de Hermanos</i>	1= Uno 2= Dos 3= Tres 4= Cuatro 5= Cinco 6= Seis 7= Siete Sin Clasificar= SC	30 (16.76%) 80 (44.69%) 29 (16.2%) 12 (6.7%) 6 (3.35%) 1 (0.55%) 1 (0.55%) 20 (11.17%)  Media = 2.314 Min – Max = 1 – 7 Mediana = 2
<i>Número de Materias Perdidas</i>	Asignaturas pérdidas durante los semestres cursados del programa académico.	Media = 11.25 Min – Max = 0 - 38 Mediana = 9
<i>Número de Materias Repetidas</i>	Una materia se pudo perder más de una vez	Media = 10.21 Min – Max = 0 - 35 Mediana = 8
<i>Puntaje en el Examen de Estado</i>	De 1 – 100 puntos	Media = 89.39 Min – Max = 44 - 100 Mediana = 93
<i>Vivienda Propia</i>	Sin Clasificar= SC Si= Si No= No	8 (4.47%) 136 (75.97%) 35 (19.55%)
<i>Trabajaba al presentar el examen de estado</i>	Si= S No= N Sin clasificar= SC	0 (0%) 159 (88.82%) 20 (11.17%)
<i>Tiempo Hasta la Deserción</i>	Semestres Académicos (16 semanas)	Media = 9.799 Min – Max = 1 - 22 Mediana = 12

<b>Variable</b>	<b>Unidades de medición /codificación</b>	<b>Resumen numérico</b>
<i>Estado del estudiante Dicotomizado</i>	Desertor= D, No Desertor= ND	78 (43.57%) 101 (56.42 %)
<i>Deserción</i>	Académica= A, No Académica= NA	34 (43.59%) 44 (56.41%)
<i>Tipos de deserción</i>	Precoz (Hasta primer semestre) Temprana (De segundo a quinto semestre) Tardía (De sexto semestre en adelante)	20 (25.64%) 43 (55.13%) 15 (19.23%)
<i>Motivo de deserción Académica (Perdió cupo de acuerdo al Artículo 80)</i>	Literal A (Quien obtenga durante cuatro (4) semestres, un promedio aritmético acumulado inferior a tres cero (3.0))= LA  Literal B (Quien teniendo un promedio aritmético acumulado inferior a tres cero (3.0), obtenga un promedio aritmético semestral inferior a dos cero (2.0))= LB  Literal C (Quien pierda una asignatura que curse como repitente siendo su promedio acumulado inferior a tres cero (3.0). Con promedio aritmético acumulado igual o superior a tres cero (3.0), la podrá cursar por tercera y última vez)= LC  Literal D (Quien pierda en un mismo periodo académico dos asignaturas que se cursen en calidad de repitente)= LD  Literal E (Quien pierda una asignatura por tercera vez)= LE	1 (1.28%)  6 (17.65%)  15 (44.12%)  6 (17.65%)  6 (17.65%)
<i>Motivo de deserción no Académica</i>	Retiro definitivo= RD  Perdió cupo Artículo 42 Literal E (Quien durante tres semestres no renueve matrícula)= NRM	37 (84.1%)  7 (15.9%)

Fuente: La Autora

Con base en lo indicado en la tabla 24, sobre los datos del programa de Ingeniería Electromecánica se afirma que la mayoría de estudiantes que ingresan a este programa son hombres (90.5%), razón por la cual la variable género es descartada como variable explicativa, se identificó que la edad promedio al momento de presentar el examen de Estado fue 17 años, el puntaje promedio obtenido en el examen de estado fue 89.39 en una escala de 0 a 100, el 75.97% cuentan con vivienda propia. Se resalta que dentro de la muestra ningún estudiante trabajaba al momento de presentar el examen de Estado, debido a ello esta variable también fue excluida como variable explicativa.

Además el 53.63% de hogares de los estudiantes tenían un ingreso entre 1 y 2 Salarios Mínimos Mensuales Legales Vigentes (SMMLV) al momento de presentar el examen de estado, otra variable relacionada con este examen es el número de hermanos ya que el 44.69% de los estudiantes tenían dos hermanos en ese momento, respecto al nivel educativo de las madres el mayor porcentaje es 39.1% quienes estudiaron hasta la básica secundaria.

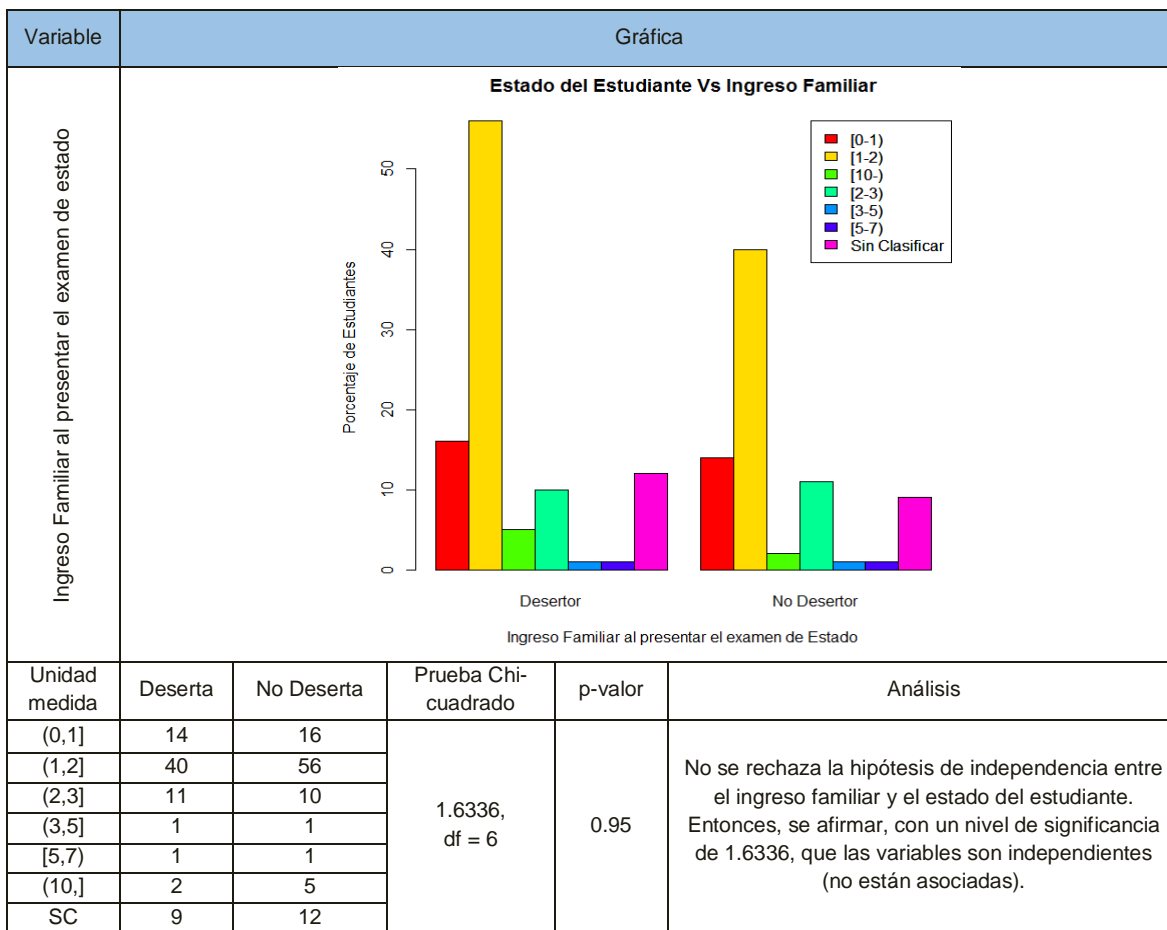
Las siguientes variables fueron tomadas hasta primer semestre de 2015, de acuerdo a lo cual se observó que el 43.57% de los estudiantes de las diez cohortes analizadas del programa Ingeniería Electromecánica desertó, durante este mismo periodo se graduó el 40.22% de los estudiantes que ingresaron, los demás estudiantes estaban matriculados y el 6.14% ya habían terminado académicamente el programa. Para el caso de los distintos tipos de deserción respecto al tiempo, la mayor concentración de los estudiantes se encuentra en la deserción temprana es decir entre el segundo y quinto semestre. Se identifica que la diferencia en la deserción académica y no académica es cercana al 13%, por ejemplo en el literal C del Artículo 80 del reglamento estudiantil (Pérdida de asignatura repetida con promedio acumulado inferior a 3.0) es la mayor causa en la deserción académica con un 44.12% del total de los desertores por motivos académicos, por otro lado la deserción por causas no académicas fue obtenida en su gran mayoría por el retiro definitivo con un 84.1% que hace referencia a los estudiantes que pasaron carta de retiro definitivo del programa.

En la descripción bivariada de las variables objeto de estudio, se analiza si las variables cualitativas medidas en los estudiantes son independientes del estado del estudiante (desertor o no desertor), para esto se usa una prueba estadística llamada la prueba Ji-cuadrado. Se encontró lo siguiente:



Tabla 25. Resumen bivariado de las variables de estudio del programa de ING

Variable	Gráfica																						
Vivienda Propia	<div><p>Estado del Estudiante Vs Vivienda Propia</p><table><thead><tr><th>Estado de vivienda del Estudiante al presentar el examen de Estado</th><th>Posee</th><th>Carece</th><th>Sin Clasificar</th></tr></thead><tbody><tr><td>Desertor</td><td>70</td><td>24</td><td>6</td></tr><tr><td>No Desertor</td><td>63</td><td>11</td><td>4</td></tr></tbody></table><p>Estado de vivienda del Estudiante al presentar el examen de Estado</p></div>					Estado de vivienda del Estudiante al presentar el examen de Estado	Posee	Carece	Sin Clasificar	Desertor	70	24	6	No Desertor	63	11	4						
Estado de vivienda del Estudiante al presentar el examen de Estado	Posee	Carece	Sin Clasificar																				
Desertor	70	24	6																				
No Desertor	63	11	4																				
Unidad medida	Desertor	No Desertor	Prueba Ji-cuadrado	p-valor	Análisis																		
Si	63	73	2.6523, df = 2	0.2655	No se rechaza la hipótesis de independencia entre las variables. Con un nivel de significancia de 2.6523 las variables son independientes (no están asociadas).																		
No	11	24																					
SC	4	4																					
Nivel Educativo de la madre	<div><p>Estado del Estudiante Vs Nivel Educativo de la Madre</p><table><thead><tr><th>Nivel educativo de la madre al presentar el examen de Estado</th><th>Básica Primaria</th><th>Básica Secundaria</th><th>Sin Clasificar</th><th>Tecnológico</th><th>Universitario</th></tr></thead><tbody><tr><td>Desertor</td><td>31</td><td>38</td><td>12</td><td>12</td><td>8</td></tr><tr><td>No Desertor</td><td>25</td><td>32</td><td>9</td><td>6</td><td>6</td></tr></tbody></table><p>Nivel educativo de la madre al presentar el examen de Estado</p></div>					Nivel educativo de la madre al presentar el examen de Estado	Básica Primaria	Básica Secundaria	Sin Clasificar	Tecnológico	Universitario	Desertor	31	38	12	12	8	No Desertor	25	32	9	6	6
Nivel educativo de la madre al presentar el examen de Estado	Básica Primaria	Básica Secundaria	Sin Clasificar	Tecnológico	Universitario																		
Desertor	31	38	12	12	8																		
No Desertor	25	32	9	6	6																		
Unidad medida	Desertor	No Desertor	Prueba Chi-cuadrado	p-valor	Análisis																		
BP	25	31	0.9315, df = 4	0.92	No se rechaza la hipótesis de independencia entre las anteriores variables. Entonces, con un nivel de significancia de 0,9315 se afirma, que las variables son independientes (no están asociadas).																		
BS	32	38																					
SC	9	12																					
TEC	6	12																					
UN	6	8																					



Fuente: La Autora

Por último, se realiza el test de Pearson que se observa en la tabla 26, para examinar la relación entre el tiempo en desertar en semestres y las variables cuantitativas usadas en este estudio.

Tabla 26. Test de correlaciones de Pearson entre tiempo en desertar y variables cuantitativas de ING

Variable cuantitativa	Tiempo hasta desertar		
	Coeficiente de Correlación	Estadístico t	p
Edad ingreso al programa	0.0728	0.615	0.541
Número de hermanos	-0.0472	-0.39	0.698
Puntaje en las pruebas de estado	0.1162	0.999	0.321

Fuente: La Autora

Se evidencia que no hay correlación significativa entre las variables listadas en la tabla 26 y el tiempo transcurrido hasta que un estudiante deserta.

## 5.2. MODELO PARA DESERCIÓN

La finalidad de esta sección es realizar la identificación de las variables o factores que están relacionados con el tiempo de la deserción de un estudiante de Ingeniería Electromecánica, se presentará la aplicación del método de Kaplan Meier para estimar la función de sobrevivencia, con ello se puede analizar la evolución de la probabilidad de la deserción con su respectivo intervalo de confianza, luego, se construirá el modelo de regresión de Cox para estimar el efecto de las variables de estudio sobre los tiempos de sobrevivencia a la deserción.

Se inicia con la definición de las características de los datos de sobrevivencia, como se expresa en la siguiente tabla:

Tabla 27. Características de los datos para el modelo de sobrevivencia

Unidad sobre la cual se registra el evento		Estudiante perteneciente a las cohortes entre 2004 y 2009 del programa de ING
El evento de interés o falla		Abandono del programa por parte del estudiante (deserción).
Variable respuesta "Tiempo hasta la deserción"		El tiempo hasta que el estudiante se va del programa ya sea por causas académicas o no académicas, la variable es discreta, ya que se mide en número de semestres cursados hasta que presenta el fenómeno de estudio.
Tiempo	Inicial del Estudio	I Semestre de 2004
	Origen del evento	Cohorte del estudiante (Primera matricula del estudiante en el programa)
	Final del Estudio	II Semestre de 2015
Censura	Tipo	Tipo I y a derecha. Los estudiantes entran al estudio en diferentes tiempos, es decir, diferentes cohortes, y el punto final del estudio es el mismo para todos. En este caso, el tiempo de censura para cada estudiante es conocido desde el momento que ingresa al estudio, con ello cada individuo tiene fijo y especificado su tiempo de censura. Se considera como censura al estudiante graduado o que continúa estudiando.
	Tiempo	Promedio de graduación en la ING, 14 semestres.

Fuente: La Autora

Se prosigue con el análisis del tiempo hasta la deserción, usando la información disponible, es decir los datos censurados y no censurados. Las probabilidades de sobrevivencia en cada intervalo así como la función de sobrevivencia se calculan con el estimador de Kaplan Meier, teniendo en cuenta que no se asumirá modelo probabilístico para el tiempo hasta la deserción y la existencia de datos censurados a derecha.

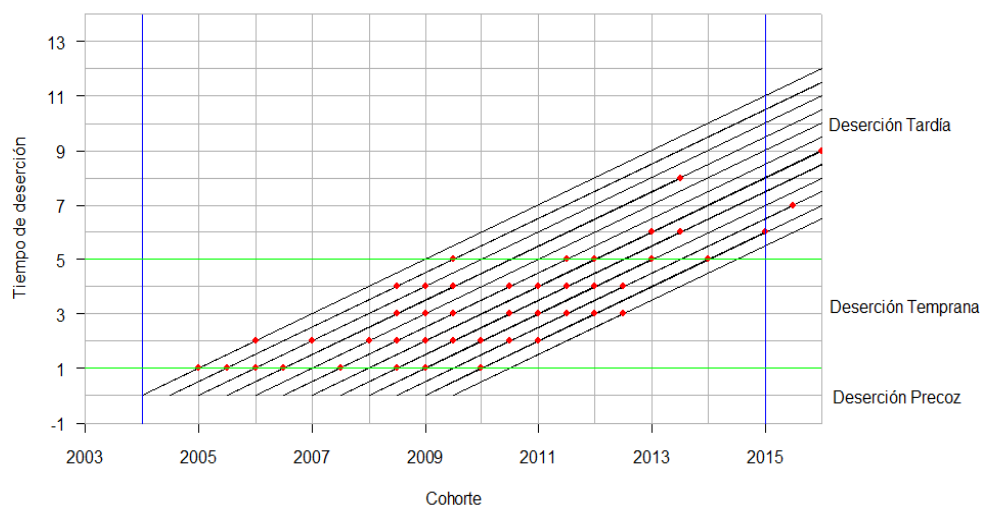
Tabla 28. Resumen porcentaje de censurados para ING

Programa	N° Total	N° de eventos	Censurado	
			N	Porcentaje
Ingeniería Electromecánica	179	77	102	56.98 %

Fuente: La Autora

En la anterior tabla, se tiene el N° total que se refiere la cantidad de estudiantes pertenecientes al programa; es el total de estudiantes, es evidente que Ingeniería Electromecánica tiene 56.98% en el porcentaje de casos censurados los casos censurados, que corresponden a los estudiantes que no presentan el evento pues se graduó o está activo en el SIRA (matriculado ó matriculado con terminación académica) al finalizar el periodo de seguimiento.

Como se inició este capítulo con las características de los datos de sobrevivencia, a continuación se tiene la representación de los estudiantes de este programa mediante el diagrama de Lexis, la sintaxis R aparece en el Anexo C, en el diagrama se refleja el tiempo calendario (semestres) en el eje horizontal y la longitud del tiempo de vida, representada por una línea a 45°. El tiempo que un individuo pasa en el estudio se representa por la altura del rayo en el eje vertical.

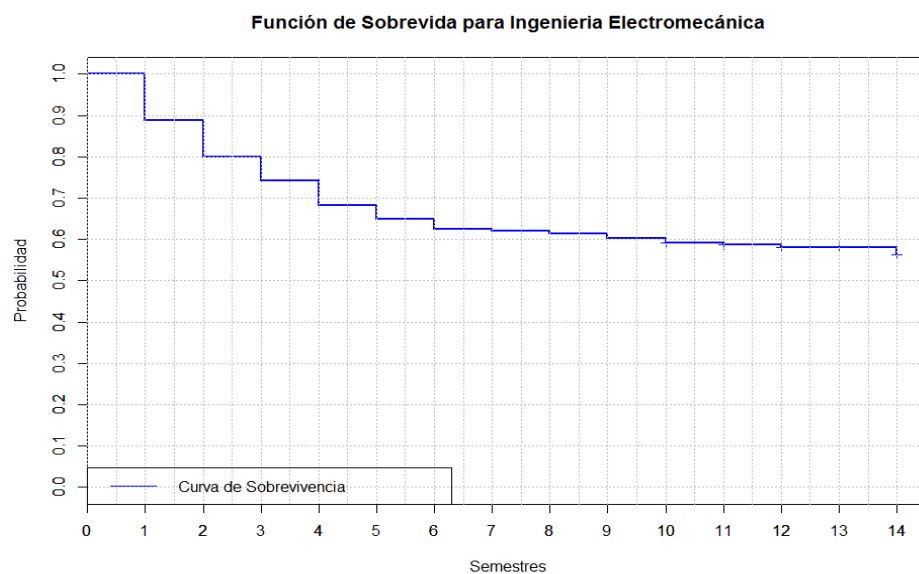


Gráfica 17. Diagrama de Lexis para ING

En el diagrama se evidencia que los estudiantes bajo estudio no tienen el mismo tiempo de origen como se mencionó en la caracterización de los datos, los puntos rojos representan aquellos estudiantes que presentaron el evento y el tiempo en que sucedió (expresado en semestres). Observando el diagrama la mayoría de estudiantes desertan precozmente, seguido de la deserción temprana y unos muy pocos se registran con deserción tardía.

### 5.2.1. Función de Sobrevida

Se realiza la estimación de la función de Sobrevida según el método de Kaplan Meier, esto se obtiene a través de R-Commander, mediante la función `survfit`, en el Anexo C se muestra la sintaxis de R. En el gráfico 18 y en la tabla 29 se observan las probabilidades de sobrevida en cada intervalo.



Gráfica 18. Función de sobrevida para ING

Se observa que la probabilidad de no desertar del programa en el primer semestre es del 88.8%; la cual va decreciendo semestre a semestre y tiende a estabilizarse en 62% en el semestre seis.

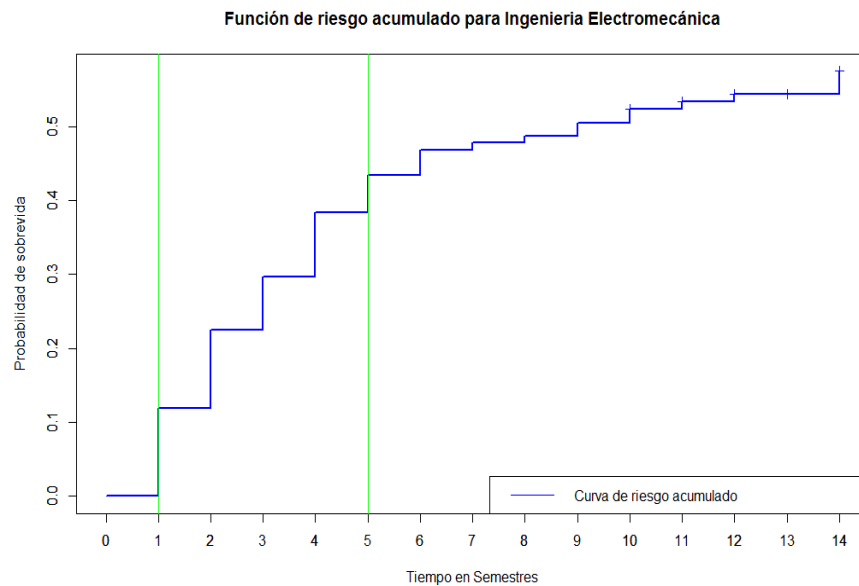
Se calcularon los cuantiles, la sintaxis R se halla en el Anexo C, a partir de estos se identifica que hay una probabilidad del 25% de desertar hasta el tercer semestre, y con una probabilidad del 50% de presentar la deserción hasta el semestre número catorce.

Tabla 29. Estimaciones función de sobrevivida para ING

Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	179	20	0.888	0.0235	0.843	0.936
2	159	16	0.799	0.0300	0.742	0.860
3	143	10	0.743	0.0327	0.682	0.810
4	133	11	0.682	0.0348	0.617	0.753
5	122	6	0.648	0.0357	0.582	0.722
6	116	4	0.626	0.0362	0.559	0.701
7	112	1	0.620	0.0363	0.553	0.695
8	111	1	0.615	0.0364	0.547	0.690
9	110	2	0.603	0.0366	0.536	0.679
10	108	2	0.592	0.0367	0.524	0.669
11	105	1	0.587	0.0368	0.519	0.663
12	99	1	0.581	0.0369	0.513	0.658
14	65	2	0.563	0.0379	0.493	0.642

Fuente: La Autora

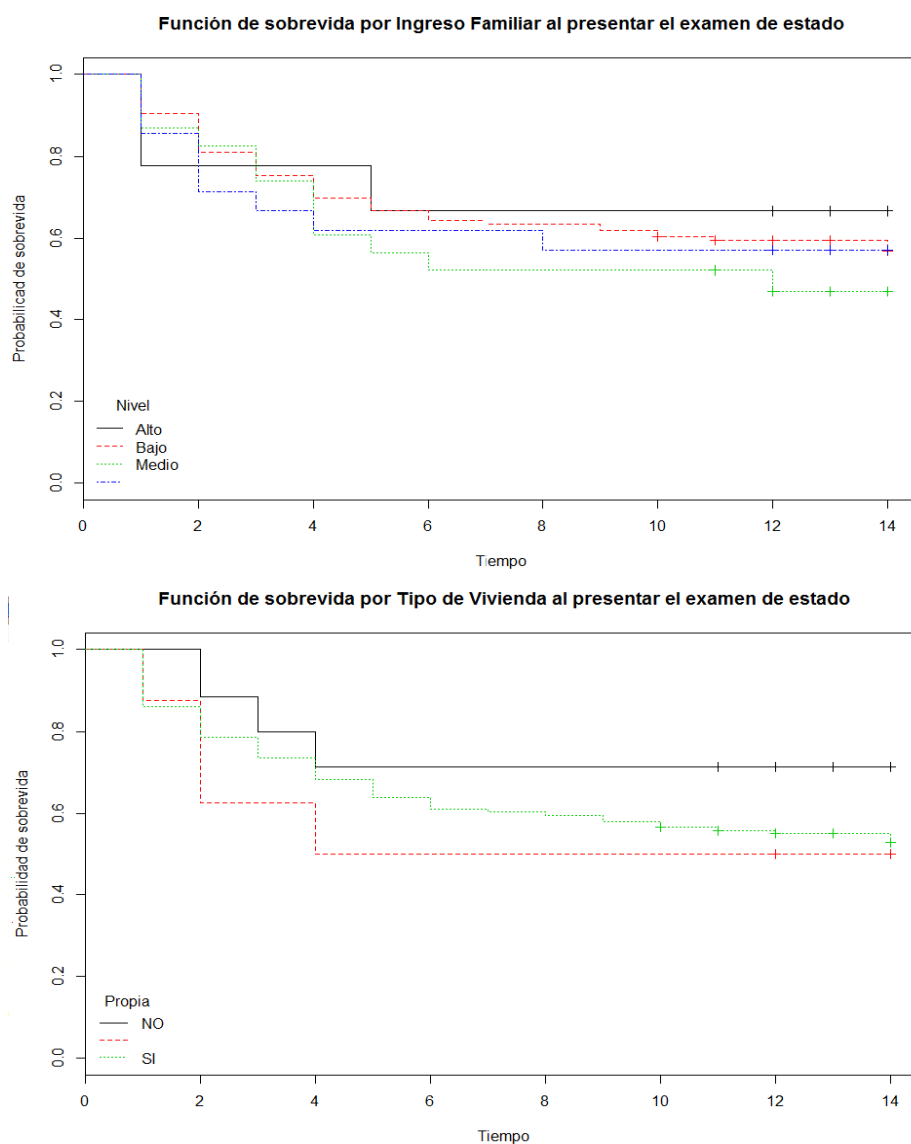
La función de riesgo, como se señaló es conocida como la tasa instantánea de mortalidad, esta función, permite tener la información del comportamiento del riesgo durante el tiempo, a continuación se presenta la correspondiente al programa de Ingeniería Electromecánica, se indican en la gráfica tanto el semestre 1 (deserción precoz), como hasta el semestre 5 (deserción temprana).



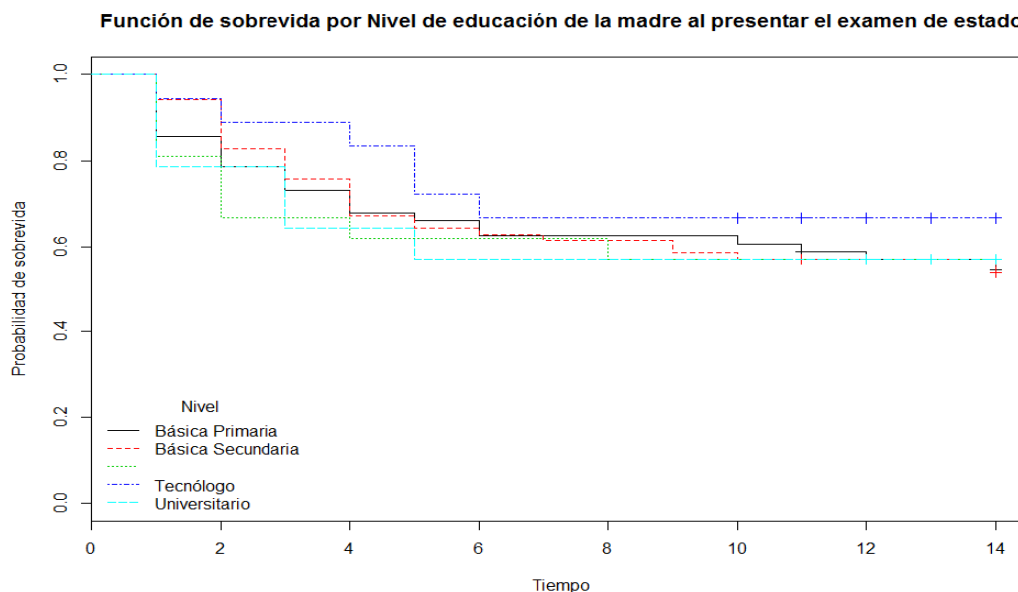
Gráfica 19. Función de riesgo acumulado para ING

A partir de la función de riesgo acumulada se puede afirmar que la probabilidad que un estudiante deserte justo al terminar el primer semestre es del 12%, es decir la probabilidad de que deserte precozmente del Programa, se evidencian que en la función de riesgo los mayores cambios están en la deserción temprana, por ejemplo, se tiene en tercer semestre que la probabilidad habiendo llegado al segundo semestre deserte en el tercero es del 7.3%.

El análisis de la función de sobrevivencia observando por separado las poblaciones de estudio según las variables categóricas, se presenta en seguida.



Gráfica 20. Otras funciones de sobrevivencia en ING



Gráfica 21. Función de sobrevivida para nivel educativo madre en ING

Se evidencia respecto a la deserción durante el tiempo, que en cuanto a la deserción precoz en el primer semestre el porcentaje de sobrevivida, presenta un gran salto entre las categorías bajo (90.5%), medio (87%) y alto (77.8%), lo cual manifiesta la variación existente en los ingresos altos y bajos; en el caso de la deserción temprana, por ejemplo en el quinto semestre las diferencias más significativas se dan entre las categorías alto (66.7%), bajo (66.7%) y medio (56.5%), y por último en la deserción tardía, se identifica en el semestre seis en los ingresos bajos y medios, a partir de esto se determina que los porcentajes de las categorías mencionadas son del 64.3% en bajos y en medios fue del 52.2% con una diferencia cercana al 12%. Al realizar el comparativo de las categorías se obtuvo un valor de  $\chi^2 = 1$  con 3 grados de libertad y un p-valor de 0.797, de acuerdo a estos resultados obtenidos se puede concluir que la sobrevivida tiene un comportamiento muy similar en las categorías de ingreso medio y bajo de esta variable, lo anterior se afirma debido a la tabla 30 y la gráfica correspondiente.

En el Anexo C se presentan las pruebas de hipótesis para las variables, nivel educativo de la madre y tenencia de propia, se identificó que para las diferentes categorías de estas variables la sobrevivencia (a la deserción) tiene el mismo comportamiento, lo cual indica poca variabilidad.



Tabla 30. Estimaciones para Ingresos familiares en ING

Ingreso=Alto					95% CI	
Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	9	2	0.778	0.139	0.549	1
5	7	1	0.667	0.157	0.420	1
Ingreso=Bajo					95% CI	
Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	126	12	0.905	0.0262	0.855	0.957
2	114	12	0.810	0.0350	0.744	0.881
3	102	7	0.754	0.0384	0.682	0.833
4	95	7	0.698	0.0409	0.623	0.783
5	88	4	0.667	0.0420	0.589	0.754
6	84	3	0.643	0.0427	0.564	0.732
7	81	1	0.635	0.0429	0.556	0.725
9	80	2	0.619	0.0433	0.540	0.710
10	78	2	0.603	0.0436	0.524	0.695
11	75	1	0.595	0.0437	0.515	0.687
14	46	2	0.569	0.0455	0.487	0.666
Ingreso=Medio					95% CI	
Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	23	3	0.870	0.0702	0.742	1.000
2	20	1	0.826	0.0790	0.685	0.996
3	19	2	0.739	0.0916	0.580	0.942
4	17	3	0.609	0.1018	0.439	0.845
5	14	1	0.565	0.1034	0.395	0.809
6	13	1	0.522	0.1042	0.353	0.772
12	10	1	0.470	0.1060	0.302	0.731
Ingreso=NA					95% CI	
Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
1	21	3	0.857	0.0764	0.720	1.000
2	18	3	0.714	0.0986	0.545	0.936
3	15	1	0.667	0.1029	0.493	0.902
4	14	1	0.619	0.1060	0.443	0.866
8	13	1	0.571	0.1080	0.395	0.828

Fuente: La Autora

### **5.2.2. Modelo de Cox – Factores que influyen en la deserción**

El modelo de Cox, mide los efectos de las variables consideradas en el estudio como explicativas del fenómeno de la deserción (edad de ingreso al Programa, nivel educativo de la madre, número de hermanos, vivienda propia, ingreso familiar al presentar el examen de estado, puntaje ICFES estandarizado), al seleccionar el modelo más parsimonioso, utilizando el método hacia adelante y con el criterio el AIC (Criterio de Información de Akaike), se encontró que ninguna variable aporta a medir la variabilidad del evento de la deserción en Ingeniería Electromecánica. Lo anterior significa que se debe continuar estudiando otras variables que puedan ser factores determinantes en el fenómeno, por ahora se descartan las consideradas en este proyecto.

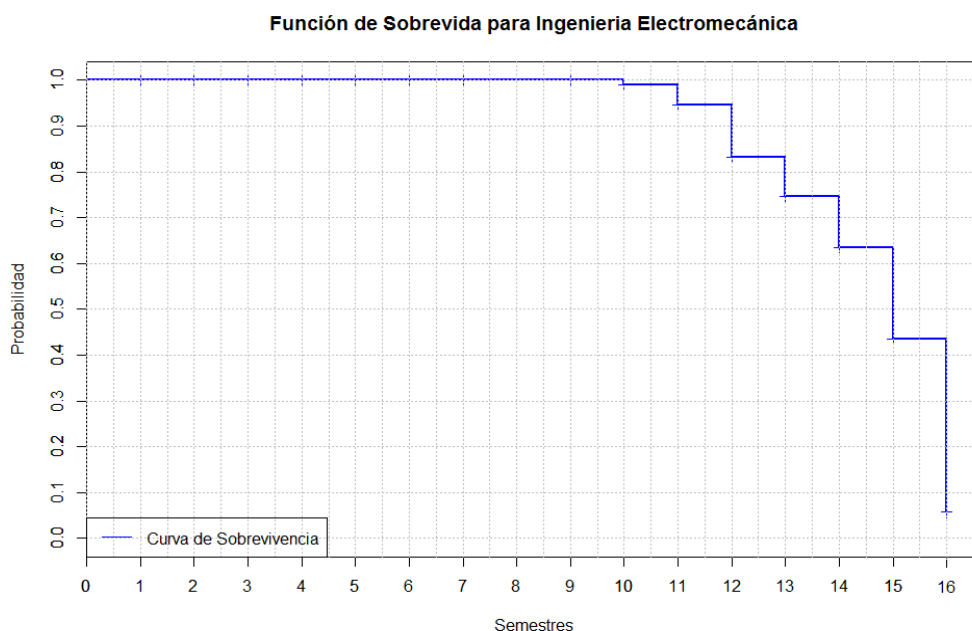
### **5.3. MODELO PARA GRADUACIÓN**

Se analizará ahora el fenómeno de la graduación, la caracterización de los datos de sobrevida corresponde a:

- Evento de interés o falla: Terminación del programa por parte del estudiante (graduación).
- La variable respuesta: Tiempo hasta que el estudiante finaliza el programa, lo cual se mide en número de semestres cursados hasta que se gradúa.
- Tipo de censura: Tipo I y a derecha, los individuos ingresan al estudio en diferentes tiempos, como ya se había mencionado, diferentes cohortes, y el punto final del estudio es el mismo para todos. Se considera como censura al estudiante que deserta (causas académicas o no académicas) o que continúa estudiando.
- Tiempo de censura: Tiempo hasta el percentil 75 que corresponde al semestre número dieciséis.

#### **5.3.1. Función de Sobrevida**

La función de sobrevida estimada para el modelo, se muestra en la gráfica 22 y su respectiva estimación en la tabla 32:



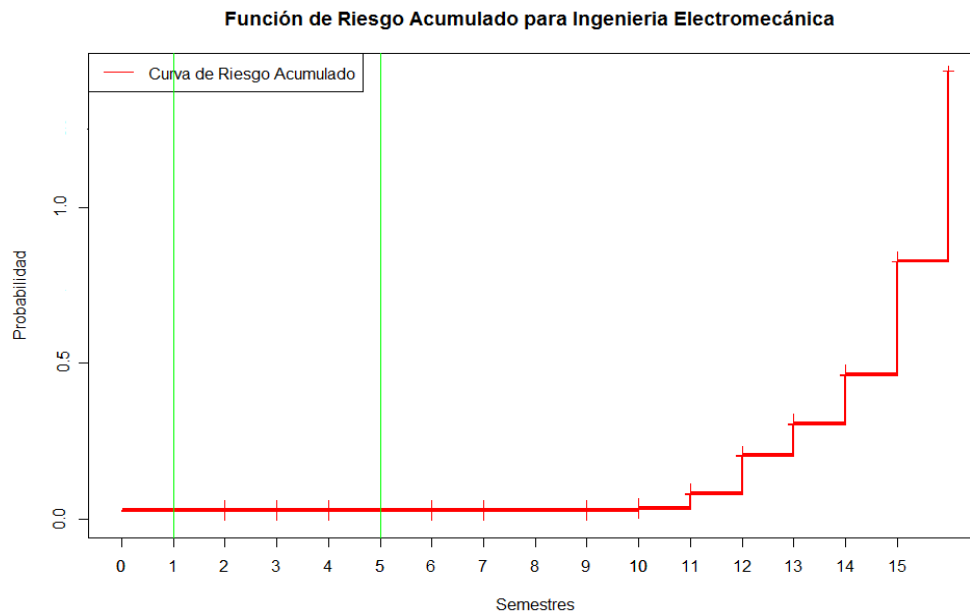
Gráfica 22. Función de sobrevida para graduación de ING

Tabla 31. Estimaciones por Kaplan Meier para graduación en ING

Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
10	93	1	0.9892	0.0107	0.969	1.000
11	90	4	0.9453	0.0238	0.900	0.993
12	84	10	0.8327	0.0394	0.759	0.914
13	68	7	0.7470	0.0468	0.661	0.845
14	53	8	0.6343	0.0541	0.537	0.750
15	35	11	0.4349	0.0621	0.329	0.575
16	23	20	0.0567	0.0316	0.019	0.169

Fuente: La Autora

Se puede mencionar que el modelo muestra que la graduación se presenta desde el semestre décimo en adelante, como era de esperarse, la probabilidad de terminar el décimo semestre y graduarse es 1%, en el semestre número doce la probabilidad de graduarse es del 18.3% y por último en el semestre quince la probabilidad de graduarse al terminarlo es 83,26%.



Gráfica 23. Función de riesgo acumulado para graduación en ING

### 5.3.2. Modelo de Cox – Factores que influyen en la graduación

El proceso se inicia con las estimaciones del modelo semiparamétrico de Cox según el tiempo hasta la graduación, los resultados del modelo con todas las variables se presentan en la tabla 33.

Tabla 32. Modelo de Cox para graduación en ING

	Coeficiente regresión	Exp (Coeficiente regresión)	Error estándar (Coeficiente regresión)	z	Pr(> z )
Edad.de.ingreso.al.programa	6,30E+01	1.065	8,64E+01	0.729	0.46580
Ingreso.Familiar[T.BJ]	-1,72E+04	0	3.957.000	-0.004	0.99653
Ingreso.Familiar[T.ME]	-16.870	0	3.957.000	-0.004	0.99660
Nivel.Educativo.Madre[T.BS]	4,02E+02	1.495	3,48E+02	1.157	0.24735
Nivel.Educativo.Madre[T.TEC]	7,80E+02	2,18E+03	4,67E+02	1.668	0.09533
Nivel.Educativo.Madre[T.UN]	-1,59E+04	0	3,96E+06	-0.004	0.99680
Numero.de.Hermanos	418	1.519	1,40E+02	2.980	0.00289 **
Puntaje	1,92E+01	1.019	1,70E+01	1.129	0.25894
Vivienda.Propia[T.SI]	-7,94E+02	452	3,54E+02	-2.244	0.02485 *

Fuente: La Autora

En la selección del modelo más parsimonioso, se utiliza el método hacia adelante y criterio el AIC (Criterio de Información de Akaike), encontrándose que el modelo óptimo está determinado por vivienda propia, número de hermanos y nivel educativo de la madre. La siguiente tabla resume el modelo óptimo:

Tabla 33. Resumen modelo óptimo para graduación en ING

	Coeficiente regresión	Exp (Coeficiente regresión)	Error estándar (Coeficiente regresión)	z	Pr(> z )	
Vivienda.Propia[T.SI]	-0.7527	0.4711	0.3367	-2.235	0.02539	*
Numero.de.Hermanos	0.4348	15.447	0.1392	3.124	0.00178	**
Nivel.Educativo.Madre[T.BS]	0.4384	15.503	0.3415	1.284	0.19924	
Nivel.Educativo.Madre[T.TEC]	0.9948	27.043	0.4155	2.394	0.01665	*
Nivel.Educativo.Madre[T.UN]	0.9275	2,53E+04	0.5486	1.690	0.09094	.
---						
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '	1
	Exp (Coeficiente regresión)	Exp (-Coeficiente regresión)	L. inferior 0.95	L. superior 0.95		
Vivienda.Propia[T.SI]	0.4711	21.227	0.2435	0.9114		
Numero.de.Hermanos	15.447	0.6474	11.759	20.291		
Nivel.Educativo.Madre[T.BS]	15.503	0.6451	0.7938	30.276		
Nivel.Educativo.Madre[T.TEC]	27.043	0.3698	11.978	61.058		
Nivel.Educativo.Madre[T.UN]	25.281	0.3956	0.8626	74.097		

Fuente: La Autora

En el anterior resumen del modelo óptimo, se observa la estimación del riesgo relativo para cada covariable del modelo en la graduación para Ingeniería Electromecánica, su intervalo de confianza y el p-valor sobre su significancia, también se tiene la inversa del riesgo relativo para los casos inferiores a 1. En seguida se explican los resultados de cada una de ellas:

Nivel educativo de la madre:

El caso del riesgo relativo estimado en el nivel educativo secundaria corresponde a 15.503 respecto a básica primaria, en otras palabras si se comparan dos estudiantes aquel que tenga una madre con nivel educativo básica primaria tendrá 0.6451 veces mayor probabilidad de graduarse respecto a un estudiante con madre de nivel educativo básica secundaria, es decir disminuye.

El riesgo relativo del nivel educativo de la madre con técnico es 27.043 respecto a básica primaria, con lo cual el riesgo de desertar se multiplica por 0.3698 veces, es decir disminuye la probabilidad de graduarse para quienes tenga madre con nivel educativo técnico al hacer la comparación con estudiantes de madres con un nivel de educación básica primaria. Esta relación es significativa.

La estimación del riesgo relativo del nivel educativo de la madre universitario es 25.281 respecto al nivel educativo básica primaria, con lo cual el riesgo de graduarse tiene una relación negativa, es decir, estudiantes con madre de nivel educativo primaria tienen 0.3956 más veces, luego disminuye la probabilidad de graduarse.

#### Vivienda Propia:

El riesgo relativo estimado de si posee vivienda propia al presentar el examen de estado respecto al no tener vivienda propia es 0.4711. Con lo cual tiene menos riesgo de graduarse los que si poseen vivienda respecto a los que no, es decir cuando un estudiante no posee vivienda se multiplica por 21.227 la probabilidad de graduarse. Esta relación si es significativa, además el p-valor es 0.02539, menor que 0.05.

#### Número de Hermanos:

La estimación del riesgo relativo respecto al número de hermanos es 15.447, se observa que el p-valor es 0.00178 (menor a 0.05), teniendo esto en cuenta y con la estimación puntual se indica que el número de hermanos incide en la terminación de los estudios, es decir, si se comparan dos estudiantes manteniendo las demás variables constantes, para aquel estudiante que tenga un hermano más, se multiplica por 0.6474 la probabilidad de terminar sus estudios, por lo tanto disminuye.

Así, se puede concluir que en el modelo ajustado para la graduación en este programa contiene ciertas variables explicativas que disminuyen la probabilidad de graduarse, nivel educativo de la madre al momento de presentar el examen de estado con las categorías básica secundaria, técnico y universitario respecto de la categoría básica primaria y tener un mayor número de hermanos, por otro lado al analizar la variable no tenencia de vivienda propia se determina que afecta de manera positiva en el tiempo de sobrevivencia, es decir debido a ello aumentan las posibilidades de graduarse.

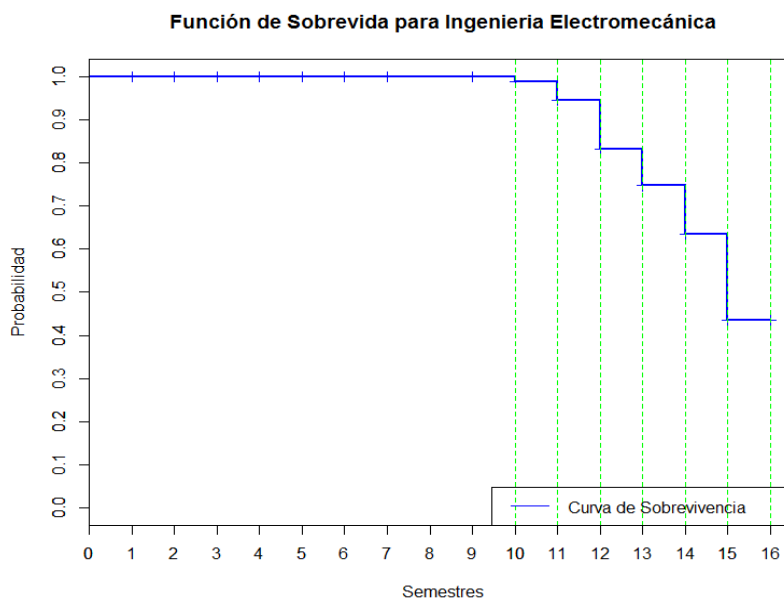
El modelo ajustado para la graduación en este programa académico resulta ser aceptable con cualquiera de los tres criterios (Likelihood, Wald y test de Puntajes) y se identifica que las tres variables seleccionadas logran explicar el 10% de la variabilidad de acuerdo al tiempo en que se presenta el evento de graduación, como se muestra en la tabla 35:

Tabla 34. Estimación del modelo de Cox de graduación en ING

Concordancia= 0.661 (se=0.055)			
Rsquare= 0.1 (max possible= 0.949 )			
Likelihood ratio test=	15.17	con 5 df,	p=0.009656
Wald test =	15.41	con 5 df,	p=0.008765
Score (logrank)=	16	con 5 df,	P=0.006838

Fuente: La Autora

Después de realizar el análisis del modelo ajustado, se determina la función de sobrevivencia, que se presenta a continuación. Se evidencia que desde la gráfica 24 y la tabla 36, la probabilidad de no graduarse en el semestre diez es del 99.1%, empieza a disminuir semestre tras semestre, un gran salto se evidencia en el semestre doce ya que la probabilidad de no graduarse a disminuido en 11.6% entonces específicamente en ese semestre se tiene una probabilidad de 83.7%, otro gran salto es notorio en el semestre catorce pues la probabilidad d no graduarse desciende al 61.5% y por último en el semestre quince se ubica cercanamente en el 44.4%.



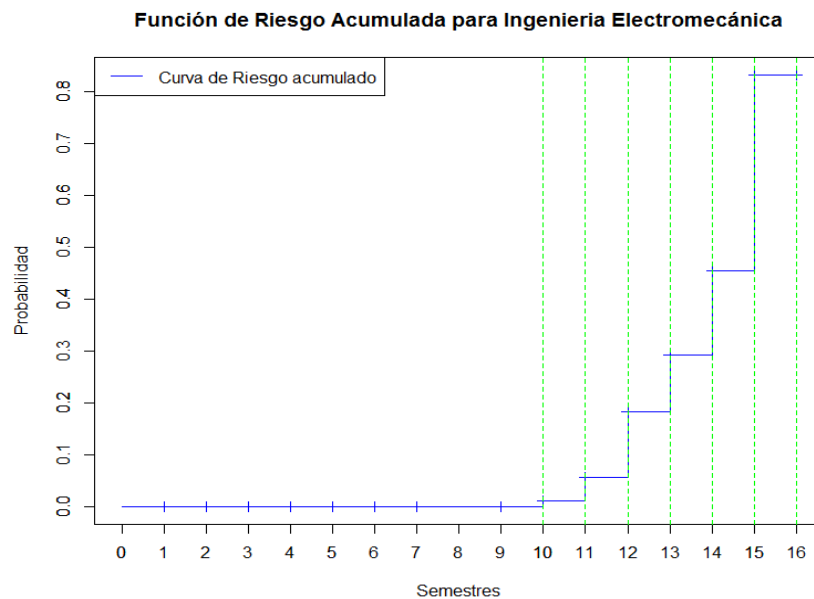
Gráfica 24. Función estimada para el modelo de graduación en ING

La función de riesgo acumulado indicada en la tabla 36 y en la gráfica 25, muestra que el riesgo de graduarse en el semestre diez es 0.009%, es decir casi improbable que eso ocurra, el semestre doce posee un riesgo de 17.79%, el cual es bajo, ya en el semestre dieciséis el riesgo de graduarse aumenta a 81.19%.

Tabla 35. Estimaciones función de sobrevivida para el modelo de graduación en ING

Semestre	Número de individuos	Número de fallas	Sobrevida	Error estándar	Límite inferior	Límite superior
10	108	1	0.991	0.00922	0.973	1.000
11	105	4	0.953	0.02052	0.914	0.994
12	99	12	0.837	0.03609	0.770	0.911
13	81	8	0.755	0.04276	0.675	0.843
14	65	12	0.615	0.05035	0.524	0.722
15	43	12	0.444	0.05558	0.347	0.567

Fuente: La Autora



Gráfica 25. Función de riesgo acumulado para graduación en ING

### 5.3.3. Evaluación del modelo de Cox

En esta sección se realiza la prueba del supuesto de riesgos proporcionales de nuestro ajuste del modelo de regresión, por medio de R-Commander, obteniendo la tabla 37.



De acuerdo a lo observado, se identifica que los p-valores son mayores al 0.05, es decir no existen evidencias significativas al 5% de que se viole el supuesto de riesgos proporcionales para ninguna de las tres covariables, ni globalmente.

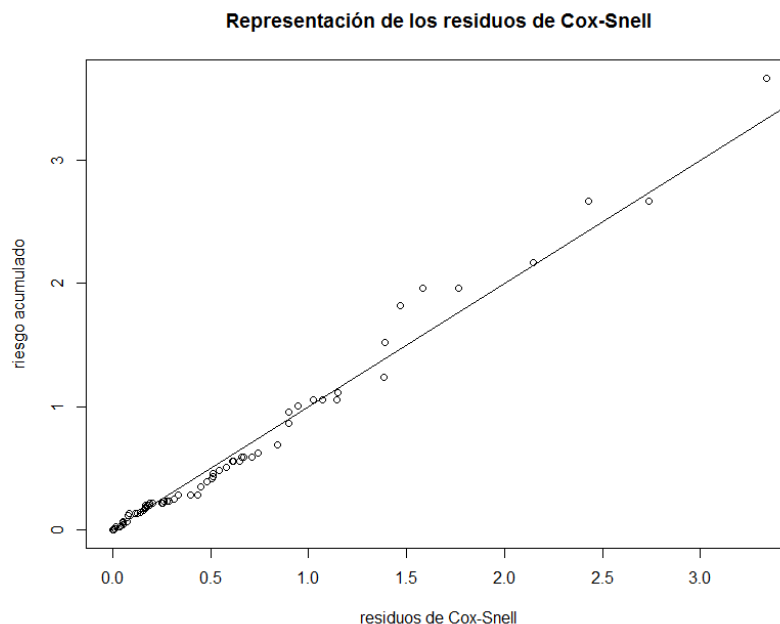
Se calculan los residuos de Cox-Snell para evaluar el ajuste del modelo de riesgos proporcionales, según lo observado en la gráfica 26, el modelo ajusta bien los datos ya que en general se acercan a la pendiente.

Tabla 36 Riesgos proporcionales en graduación de ING

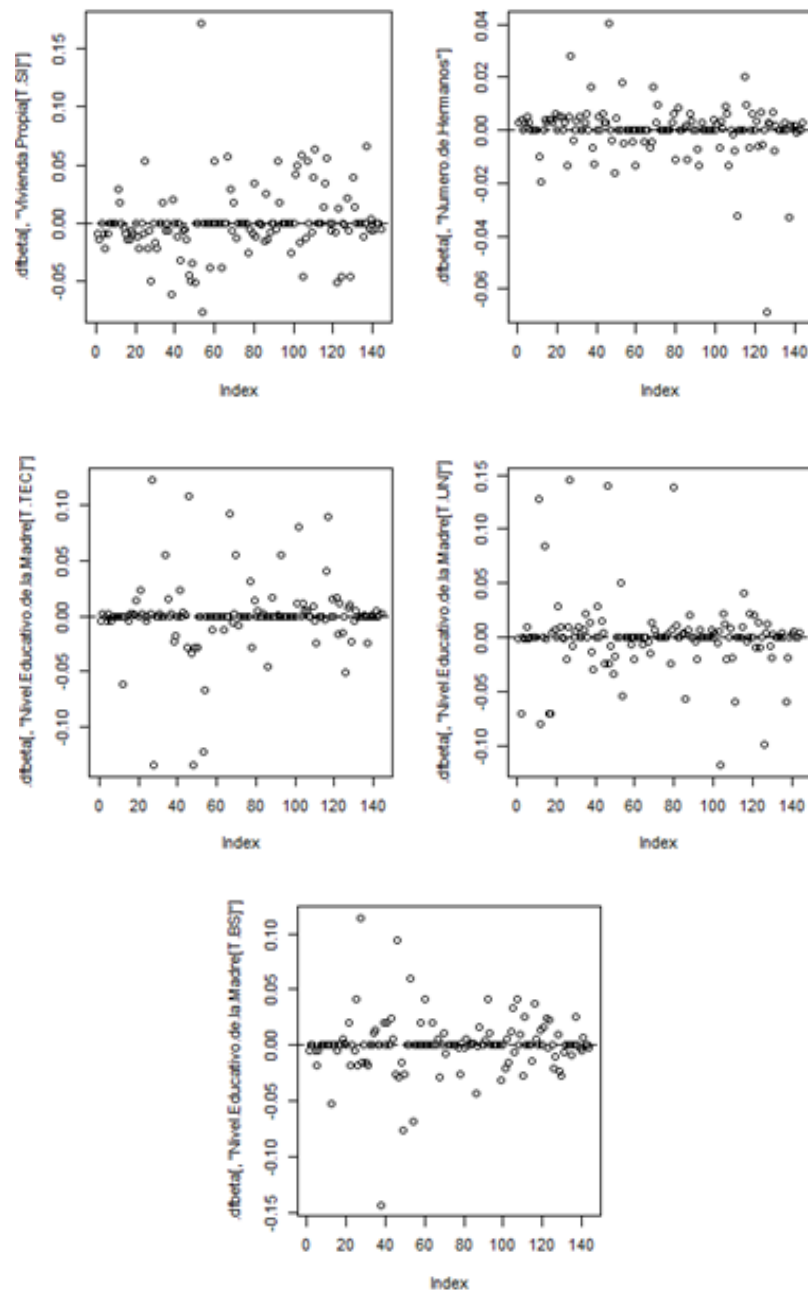
	rho	chisq	P
Vivienda.Propia[T.SI]	-0.0921	0.5973	0.440
Numero.de.Hermanos	0.1694	14.766	0.224
Nivel.Educativo.de.la.Madre[T.BS]	-0.1503	11.681	0.280
Nivel.Educativo.de.la.Madre[T.TEC]	-0.1238	0.9472	0.330
Nivel.Educativo.de.la.Madre[T.UN]	0.0416	0.0939	0.759
GLOBAL	NA	54.306	0.366

Fuente: La Autora

La comprobación de la influencia sobre cada observación en el modelo (Residuos dfbeta), por cada covariable se representa según la observación (tiempo de falla) por el cambio de escala (dividiendo por el error estándar):



Gráfica 26. Residuos de Cox-Snell para graduación en ING



Gráfica 27. Residuos de Cox-Snell para graduación en ING

Ahora en la evaluación del modelo respecto al fenómeno de la graduación en el programa de Ingeniería Electromecánica, los residuos de la anterior figura están centrados cerca al origen para el caso de las variables explicativas vivienda propia, número de hermanos y nivel educativo de la madre. Aunque se evidencia la presencia de algunos datos demasiados alejados del origen en cada covariable.

En el programa de Ingeniería Electromecánica se identificó que el 43.57% de los estudiantes que ingresan desertan, cifra que según el MEN (2014) está por debajo del nivel universitario para Colombia que fue del 44.9% para el año 2013, respecto a la de Boyacá para ese mismo año es ligeramente superior pues está en 43.2% al comparar con el área básica de conocimiento del Programa— Ingeniería, arquitectura, urbanismo y Afines - también es inferior ya que se encuentra en 50,7%. Es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 44, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

Se puede afirmar que la mayoría de estudiantes que ingresan a este programa son hombres (90.5%), razón por la cual la variable género es descartada como variable explicativa, el puntaje promedio obtenido en el examen de estado fue 89.39 en una escala de 0 a 100, el 75.97% contaban vivienda propia. Se resalta que dentro de la muestra ningún estudiante trabajaba, debido a ello esta variable también fue excluida como variable explicativa. Además el 53.63% de hogares de los estudiantes tenían un ingreso entre 1 y 2 Salarios Mínimos Mensuales Legales Vigentes (SMMLV) al momento de presentar el examen de estado, otra variable relacionada con este examen es el número de hermanos ya que el 44.69% de los estudiantes tenían dos hermanos en ese momento, respecto al nivel educativo de las madres el mayor porcentaje es 39.1% correspondiente a básica secundaria. El literal C del Artículo 80 del reglamento estudiantil es la mayor causa en la deserción académica con un 44.12% del total de los desertores por motivos académicos, por otro lado la deserción por causas no académicas fue obtenida en su gran mayoría por el retiro definitivo con un 84.1% que hace referencia a los estudiantes que pasaron carta de retiro definitivo del programa.

Para el programa, la posibilidad de que un estudiante se vaya en primer semestre es del 12%, donde el periodo crítico, es decir con mayor intensidad es de manera temprana (del primero al quinto semestre). Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas.

Para ING, el modelo de sobrevivencia para la deserción permitió identificar que a partir de las variables consideradas en este estudio no se puede detectar aquellas que inciden en tal fenómeno, por lo cual se recomienda analizar otras variables para un próximo estudio.

Para la graduación los factores que disminuyen son la tenencia de vivienda, es decir los estudiantes que carecen de vivienda tiene una mayor probabilidad de graduarse, estudiantes con madres de nivel educativo de básica primaria al momento de presentar el examen de estado respecto a un estudiante con madre de otro nivel educativo y un caso contrario es el número de hermanos pues a mayor número de hermanos menor probabilidad de graduarse.

## 6. CONCLUSIONES Y RECOMENDACIONES

Como conclusión principal del programa de Diseño Industrial se tiene que el 60% de los estudiantes que ingresan desertan, cifra que según el MEN (2014) está por encima del nivel universitario para Colombia que fue del 44.9% para el año 2013, la de Boyacá para ese mismo año que está en 43.2% e incluso la del área básica de conocimiento del Programa – Ingeniería, arquitectura, urbanismo y Afines - que se encuentra en 50,7%. Este hecho es preocupante ya que de 100 estudiantes que ingresan al Programa lo abandonan 60, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables. Para el caso de DI se identificó que el 26.25% de los estudiantes se gradúan.

La mayoría desertan por causas no académicas es decir que se dejan de matricular tres semestres consecutivos o solicitan “retiro definitivo” (Artículo 42 / Acuerdo 130 de 1998). Desertan en su mayoría los hombres, características sobresalientes de los estudiantes que ingresaron a Diseño Industrial son la edad promedio de ingreso (17 años), el 65% de los estudiantes contaban con vivienda propia, el 84% no trabajaban al momento de presentar el examen de estado, el puntaje promedio del examen de estado fue de 67.6, el 35,6% de los estudiantes tenían madres con nivel educativo de básica primaria y el 44,5% de los estudiantes tenían dos hermanos al momento de presentar el examen de Estado.

Para el programa de DI, la posibilidad de que un estudiante se vaya en primer semestre es del 19%, donde el periodo crítico con mayor intensidad es de manera temprana (del primero al tercer semestre). Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas.

Al determinar qué variables están asociadas con el estado del estudiante (desertor o no desertor) se determinó que sólo ocurre con la variable “trabajaba a la hora de presentar el examen de estado también llamado prueba SABER 11” y que el semestre en que deserta y el puntaje en las pruebas de estado están correlacionados positivamente, es decir, a medida que aumenta el puntaje también aumenta el tiempo hasta que un estudiante se va del programa.

El factor que aumenta el riesgo de desertar en DI es trabajar al momento de presentar su examen de estado, lo cual se puede atribuir a la falta de tiempo con

las actividades académicas, es decir que en la medida que un estudiante no trabaje se reduce el riesgo de deserción, se recomienda a la Universidad y al programa crear políticas de patrocinio y apoyo a aquellos estudiantes con esta característica. Se disminuye el riesgo de desertar al presentar mayor edad al ingresar al programa, lo cual puede evidenciarse debido al tema vocacional en los estudiantes que ingresan muy jóvenes a DI.

Por el contrario, para la graduación el factor que disminuye esta propensión es mayor número de asignaturas perdidas, luego un estudiante presenta menor probabilidad de graduarse. Además a partir del semestre número catorce es más probable que un estudiante se gradúe en el 14, hecho que es preocupante ya que gasta cuatro semestres más de lo previsto para obtener el título profesional y posiblemente entre más tiempo transcurra será difícil culminar con el proyecto educativo. Por lo anterior se sugirieron actividades para aumentar los niveles de graduación inmediatamente se culmine con la respectiva carga académica, por ejemplo la asignación de temas de grado en semestres inferiores, vinculación inmediata con algún programa de posgrado según la normatividad de la Universidad, o en la pasantía involúcralos con proyectos productivos.

En el programa de Ingeniería Electromecánica se identificó que el 43.57% de los estudiantes que ingresan desertan, cifra que según el MEN (2014) está por debajo del nivel universitario para Colombia que fue del 44.9% para el año 2013, respecto a la de Boyacá para ese mismo año es ligeramente superior pues está en 43.2% al comparar con el área básica de conocimiento del Programa– Ingeniería, arquitectura, urbanismo y Afines - también es inferior ya que se encuentra en 50,7%. De 100 estudiantes que ingresan al Programa lo abandonan 44, constituyéndose en una frustración personal, a nivel institucional las pérdidas económicas son altas por las inversiones y en lo social los impactos son considerables.

Se puede afirmar que la mayoría de estudiantes que ingresan a este programa son hombres (90.5%), el puntaje promedio obtenido en el examen de estado fue 89.39 en una escala de 0 a 100, el 75.97% contaban vivienda propia. Se resalta que dentro de la muestra ningún estudiante trabajaba. Además el 53.63% de hogares de los estudiantes tenían un ingreso entre 1 y 2 Salarios Mínimos Mensuales Legales Vigentes (SMMLV) al momento de presentar el examen de estado, otra variable relacionada con este examen es el número de hermanos, el 44.69% de los estudiantes tenían dos hermanos en ese momento, respecto al nivel educativo de las madres el mayor porcentaje es 39.1% correspondiente a básica secundaria. El literal C del Artículo 80 del reglamento estudiantil es la mayor causa en la deserción

académica con un 44.12% del total de los desertores por motivos académicos, por otro lado la deserción por causas no académicas fue obtenida en su gran mayoría por el retiro definitivo con un 84.1% que hace referencia a los estudiantes que pasaron carta de retiro definitivo del programa.

El 40.22% de los estudiantes que ingresaron entre el 2004 y 2009 se graduaron y el 43.57% desistieron de estudiar Ingeniería Electromecánica, predominando la deserción temprana (entre el segundo y quinto semestre), el literal C del Artículo 80 del reglamento estudiantil fue el mayor motivo de deserción académica y la deserción por causas no académicas tiene como mayoría a los estudiantes que se retiran voluntariamente del Programa.

Para el programa, la posibilidad de que un estudiante se vaya en primer semestre es del 12%, donde el periodo crítico, es decir con mayor intensidad es de manera temprana (del primero al quinto semestre). Este fenómeno se atribuye muchas veces a la dificultad de adaptarse a la vida universitaria y en este caso quizá al tema vocacional, en donde los jóvenes no tienen claridad sobre su proyecto de vida como parece ser el caso ya que su retiro es por causas no académicas.

Para ING, el modelo de sobrevida para la deserción permitió identificar que a partir de las variables consideradas en este estudio no se puede detectar aquellas que inciden en tal fenómeno, por lo cual se recomienda analizar otras variables para un próximo estudio. Para la graduación los factores que disminuyen son la tenencia de vivienda, es decir los estudiantes que carecen de vivienda tiene una mayor probabilidad de graduarse, estudiantes con madres de nivel educativo de básica primaria al momento de presentar el examen de estado respecto a un estudiante con madre de otro nivel educativo y un caso contrario es el número de hermanos pues a mayor número de hermanos menor probabilidad de graduarse.

A partir de los resultados obtenidos se puede apreciar que los factores estudiados que afectan la deserción y la graduación (excepto el número de materias repetidas) son exógenos, difícilmente pueden ser cambiados por la universidad. Es decir los riesgos se asocian más con cuestiones de tipo estructural de nuestra sociedad. Así las cosas, es importante que la UPTC encamine sus esfuerzos a mecanismos alternativos de estudio que le permita a estudiantes con ciertas características “desfavorables” reducir los riesgos de deserción y aumentar sus posibilidades de graduación.

Es importante analizar la graduación en ambos programas de estudio pues según lo encontrado la probabilidad de terminar décimo semestre y no graduarse es muy alta, es después del semestre doce cuando inicia el aumento de la probabilidad de graduarse.

Se recomienda construir a futuro nuevos modelos de sobrevivencia a partir de la recolección de información a través del instrumento, el cual está disponible en el "Anexo D", y que tiene como propósito recolectar la información sobre aspectos importantes que no fueron abordados en este proyecto.

En este estudio se usó para construir la función de sobrevivencia técnicas no paramétricas, sería conveniente en los trabajos posteriores asumir una distribución de probabilidad para las variables tiempo hasta la deserción y graduación.



## BIBLIOGRAFÍA

Barrera, M. (2008). *Análisis de supervivencia aplicado al problema de la deserción estudiantil en la Universidad Tecnológica de Pereira*. Proyecto de maestría.

Borges, R. (2005). *Análisis de Supervivencia Utilizando el Lenguaje R*. Paipa, Boyacá, Colombia.

Castano, E.; Durán, D.; Franco, J.; Gallón, S.; Gómez, K.; Guzmán, C.; Vásquez, J.; (2009). *Deserción estudiantil en la educación superior colombiana, metodología de seguimiento, diagnóstico y elementos para su prevención*. Obtenido de:

[http://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-254702\\_libro\\_desercion.pdf](http://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-254702_libro_desercion.pdf). [12 de Febrero de 2015]

Colosimo, Enricon A. y Giolo Ruíz, Suely.(2006) Conceptos básicos y ejemplos, técnicas no paramétricas y modelos probabilísticos. En: *Análise de sobrevivência aplicada*. 1 Ed. Edição: 1ª . Edgard Blucher, 1997. P. 6-64.

Correa, C. A., & Páramo, G. J. (1999). Deserción Estudiantil Universitaria. Conceptualización. *Universidad Eafit*, páginas: 4 y 5.

Cox, D.R. (1972) "*Regression models and life tables*" en Journal of the Royal Statistic Society, series B, vol 34, num 2, pp 187 - 220.

Díaz, L. G. (2015). Cursillo: Análisis estadístico de datos "*Tiempo para un evento*". *III Encuentro Internacional de Matemáticas*. Duitama.

Doménech, J. (1992). *Una aplicación del análisis de la supervivencia en ciencias de la salud*.

García Hinojosa, C. P. (2014). Estudio de análisis de supervivencia.

Giovangnoli, P. I. (2002). *Determinantes de la deserción y graduación universitaria: Una Aplicación de los modelos de duración*. Documento de trabajo, núm 37, Argentina, Universidad Nacional de la Plata.

Godoy, A. M. (2009). *Introducción al Análisis de Supervivencia con R*. Universidad Autónoma de México.

Higueta Rico, Alberto D. *Caracterización de la deserción estudiantil en la Universidad Nacional de Colombia sede Medellín*. Medellín: Colombia. Universidad Nacional de Colombia sede Medellín. 2006. p. 8-13 [3 de Marzo de 2015]. Obtenido de:  
<http://www.medellin.unal.edu.co/dirplanea/documentos/EstudioDesercionUnalMed.pdf>

Lerma, H. D. (2009). *Metodología de la Investigación: Propuesta, Anteproyecto, Proyecto*. Bogotá: Eco Ediciones.

Lininger, C. A., & Warwick, D. P. (1978). *La Encuesta por Muestreo: Teoría y Práctica*. México: Continental, S.A .

Lopera, C. (2008). "*Determinantes de la deserción Universitaria en la facultad de economía*" Universidad del Rosario, en borradores de Investigación, pp. 1-25.

Lopez, M. (2011). "Comprobación de dos modelos de regresión en fiabilidad" Universidad de Granada.

Ministerio de Educación Nacional. (30 de Octubre de 2007). *Proyectos institucionales para disminuir la deserción en educación superior*. Obtenido de <http://www.mineduacion.gov.co/1621/article-133057.html>

Ministerio de Educación Nacional. (2010). *Educación de calidad el camino hacia la prosperidad*. Obtenido de:  
<http://www.mineduacion.gov.co/cvn/1665/w3-article-254383.html>

Osorio, A., Bolancé, C., & Castillo, M. (2009) "*Modelos de deserción de supervivencia universitaria: Una aplicación de los modelos de supervivencia*", en Revista Iberoamericana de Educación Superior (ries), México, unam-iisue/Universia, Volumen. III, número. 6, páginas. 31-57. Obtenido de: <http://ries.universia.net/index.php/ries/article/view/97> [12 de Diciembre de 2014].

Ospina, D. B. (2001). *Introducción al Muestreo*. Bogotá.D.C: UNIBIBLOS.

Palacio, A. (2010). *La comprensión clásica del suicidio. De Émile Durkheim a nuestros días*. Revista Afectio Societatis, Volumen VII, número 12.

Pérez, C. (2000). *Técnicas de Muestreo Estadístico*. Madrid: Alfaomega.  
Universidad Nacional de Colombia. (Mayo de 2006). *Universidad Nacional de Colombia Sección Medellín*.

**ANEXOS**  
**ANEXO A.CÓDIGOS ESTUDIANTES MUESTRA DE INGENIERÍA**  
**ELECTROMECAÁNICA**

4244	59191	69225	200720671	200820235
44176	59200	69226	200720674	200820267
44188	59203	69231	200720767	200820319
44191	59206	69234	200720780	200820349
44202	59208	69236	200720846	200820409
44203	59209	69239	200720864	200820414
44210	59211	69242	200720899	200820454
44262	59212	69244	200721007	200820502
44263	59213	200710371	200721927	200820840
44264	59217	200711169	200722282	200821586
49065	59220	200711236	200722453	200821658
49068	59221	200711295	200810020	200821707
49073	59223	200711373	200810050	200821903
49076	59225	200711381	200810218	200910003
49080	59229	200711467	200810289	200910006
49085	59230	200711480	200810454	200910143
49088	59254	200711507	200810546	200910148
49094	64065	200711509	200810548	200910152
49098	64066	200711530	200810549	200910283
49099	64072	200711556	200810551	200910351
49100	64074	200711668	200810554	200910352
49259	64077	200711703	200810566	200910436
49270	64080	200711996	200810611	200910586
54073	64082	200720051	200810620	200910601
54074	64091	200720141	200810628	200910770
54085	64094	200720164	200810695	200910973
54086	64102	200720187	200820018	200910980
54088	64109	200720220	200820022	200920752
54089	64110	200720227	200820032	200920766
54091	64277	200720248	200820050	200920814
54097	69202	200720368	200820095	200920983
54099	69203	200720432	200820113	200921246
54109	69204	200720495	200820142	200921256
54273	69208	200720537	200820152	200921361
54276	69216	200720592	200820185	200921457
59190	69224	200720660	200820218	

## ANEXO B. CÓDIGOS ESTUDIANTES MUESTRA DE DISEÑO INDUSTRIAL

39266	54280	69000	200720675	200821589
44011	54281	69004	200720774	200821630
44013	59002	69005	200720793	200821769
44019	59005	69006	200720811	200910377
44025	59009	69009	200720949	200910574
44030	59013	69012	200721362	200910599
44037	59014	69013	200810012	200910600
44241	59021	69017	200810019	200910618
49003	59022	69251	200810228	200910627
49004	59023	69254	200810306	200910801
49006	59025	69257	200810552	200910818
49009	59029	200711073	200810557	200910850
49012	59031	200711111	200810565	200910940
49019	59032	200711131	200810605	200910961
49020	59033	200711337	200810610	200910968
49022	59037	200711455	200810618	200910970
49023	59039	200711479	200810772	200911046
49025	59041	200711517	200810793	200911731
49026	59044	200711518	200811708	200920713
49028	59260	200711535	200811805	200921108
49031	64002	200711566	200811931	200921180
49033	64004	200711579	200820107	200921219
49034	64007	200711610	200820162	200921292
54004	64013	200711709	200820165	200921299
54008	64020	200712060	200820343	200921334
54012	64022	200712143	200820467	200921379
54015	64023	200712750	200820473	200921384
54018	64024	200720102	200820565	200921397
54020	64025	200720120	200820627	200921398
54027	64027	200720125	200820629	200921444
54028	64028	200720475	200820666	200921451
54032	64029	200720489	200821025	200922256
54033	64039	200720577	200821525	200922834
54036	64041	200720594	200821575	200923440

## ANEXO C. SENTENCIAS DE R

```
#####  
#####DISEÑO INDUSTRIAL#####  
#####
```

```
### Pasamos la base de datos de xls de Diseño Industrial al R-Commander###  
Datos_DI <- sqlQuery(channel = 1, select * from [DI$])  
names(Datos_DI) <- make.names(names(Datos_DI))
```

```
#####DESCRIPCIÓN DE LOS DATOS DISEÑO INDUSTRIAL#####
```

```
###Resumen Univariado###  
summary(Datos_DI)
```

```
###Resumen Bivariado###  
##Variables categóricas##
```

```
#Variables: Género y Desertar#  
.Table <- xtabs(~Genero+Deserto, data=Datos_DI)  
Género <- xtabs(~Genero+Deserto, data=Datos_DI)  
barplot(Género,scale="percentages",main="Estado del Estudiante Vs  
Género",xlab="Estado del Estudiante",ylab="Porcentaje de  
Estudiantes",names.arg=c("Desertor","No Desertor"),  
beside=TRUE,col=topo.colors(2))  
legend("topleft",c("Femenino","Masculino"),fill=topo.colors(2))
```

```
#Variables: Vivienda propia y Desertar#  
.Table <- xtabs(~Vivienda.Propia+Deserto, data=Datos_DI)  
Vivienda <- xtabs(~Vivienda.Propia+Deserto, data=Datos_DI)  
barplot(Vivienda,scale="percentages",main="Estado del Estudiante Vs Vivienda  
Propia",xlab="Estado de vivienda del Estudiante al presentar el examen de  
Estado",ylab="Porcentaje de Estudiantes",names.arg=c("Desertor","No Desertor"),  
beside=TRUE,col=heat.colors(3))  
legend("topleft",c("Posee","Carece","Sin Clasificar"),fill=heat.colors(3))
```

```
#Variables:Trabajaba al presentar examen de estado y Desertar#
```

```
.Table <- xtabs(~Trabajaba.al.Presentar.el.Examen.de.Estado+Deserto,
data=Datos_DI)
Trabajo <- xtabs(~Trabajaba.al.Presentar.el.Examen.de.Estado+Deserto,
data=Datos_DI)
barplot(Trabajo,scale="percentages",main="Estado del Estudiante Vs
Trabajaba",xlab="Estado de trabajo del Estudiante al presentar el examen de
Estado",ylab="Porcentaje de Estudiantes",names.arg=c("Desertor","No
Desertor"),beside=TRUE,col=topo.colors(3))
legend("topright",c("No","Sin Clasificar","Si"),fill=topo.colors(3))
```

#Variables:Nivel educativo de la madre al presentar examen de estado y Desertar#

```
.Table <- xtabs(~Nivel.Educativo.de.la.Madre+Deserto, data=Datos_DI)
Nivel.educa <- xtabs(~Nivel.Educativo.de.la.Madre+Deserto, data=Datos_DI)
barplot(Nivel.educa,scale="percentages",main="Estado del Estudiante Vs Nivel
Educativo de la Madre",xlab="Nivel educativo de la madre al presentar el examen
de Estado",ylab="Porcentaje de Estudiantes",names.arg=c("Desertor","No
Desertor"),beside=TRUE,col=topo.colors(5))
legend(4,35,c("Básica Primaria","Básica Secundaria","Sin
Clasificar","Tecnológico","Universitario"),cex=0.95,fill=topo.colors(5))
```

#Variables:Ingreso familiar al presentar examen de estado y Desertar#

```
.Table <- xtabs(~Ingreso.Familiar.al.Presentar.el.Examen.de.Estado+Deserto,
data=Datos_DI)
Ingreso <- xtabs(~Ingreso.Familiar.al.Presentar.el.Examen.de.Estado+Deserto,
data=Datos_DI)
barplot(Ingreso,scale="percentages",main="Estado del Estudiante Vs Ingreso
Familiar",xlab="Ingreso Familiar al presentar el examen de
Estado",ylab="Porcentaje de Estudiantes",names.arg=c("Desertor","No
Desertor"),beside=TRUE,col=rainbow(6))
legend("topright",c("[0-1)","[1-2)","[10-)","[2-3)","[3-5)","Sin
Clasificar"),fill=rainbow(6))
```

##Correlación de Pearson de variables continuas en Diseño Industrial##

```
#Variables: Tiempo hasta desertar, Edad de Presentación del Examen de Estado#
with(Datos_DI,cor.test(Tiempo.hasta.que.deserto,Edad.de.Presentación.del.Exam
en.de.Estado,alternative="two.sided",method="pearson",exact=NULL,conf.level=0.
95))
```

```
#Variables: Tiempo hasta desertar, Número de hermanos#
with(Datos_DI,cor.test(Tiempo.hasta.que.deserto,Numero.de.Hermanos,alternative
="two.sided",method="pearson",exact=NULL,conf.level=0.95))
```

```
#Variables: Tiempo hasta desertar, Puntaje#
with(Datos_DI,cor.test(Tiempo.hasta.que.deserto,Puntaje,alternative="two.sided",
method="pearson",exact=NULL,conf.level=0.95))
```

```
#Variables: Tiempo hasta desertar, Materias perdidas#
with(Datos_DI,cor.test(Tiempo.hasta.que.deserto,Materias.Perdidas,alternative="t
wo.sided",method="pearson",exact=NULL,conf.level=0.95))
```

```
#Variables: Tiempo hasta desertar, Materias repetidas#
with(Datos_DI,cor.test(Tiempo.hasta.que.deserto,Materias.Repetidas,alternative="t
wo.sided",method="pearson",exact=NULL,conf.level=0.95))
```

```
#####MODELO DE DESERCIÓN EN DISEÑO INDUSTRIAL#####
```

```
#Definición datos sobrevivientes#
attr(Datos_DI, "time1") <- "Tiempo"
attr(Datos_DI, "event") <- "censurados"
```

```
#Diagrama de LEXIS#
LL <- Lexis.diagram( age=c(-1,15), date=c(2003,2015), entry.age= Edad.ingreso,
exit.age= Edad.salida, birth.date=Cohorte, fail=(censurados), lwd.life=1,
cex.fail=0.8, col.fail=c("blue","red"), alab="Tiempo de deserción", dlab="Cohorte",
int=c(1,1),data=Datos_DI)
abline( v=c(2004,2015), col="blue")
abline( h=1, col="green")
abline( h=5, col="green")
```

```
#Identifica personas según entrada y salida#
text( LL$exit.date, LL$exit.age, paste(1:nrow(LL)), col="Black", font=10, adj=c(0,1))
```

```
#Estimado la función de sobrevida por kaplan- meier#
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~ 1, conf.type="log",
conf.int=0.95, type="kaplan-meier", error="greenwood", data=Datos_DI)
```



.Survfit

#Gráfico Función de sobrevida con ajustes gráficos#

```
plot(.Survfit, conf.int=FALSE,  
mark.time=TRUE,xlab="Semestres",ylab="Probabilidad",main="Función de  
Sobrevida para Diseño Industrial",lwd=2,col="red")  
box(lwd=1, col ="black")  
axis(1, seq(0,14,1))  
axis(2, seq(0,1,.1))  
abline(h = seq(0,1,.1), v =seq(0,14,.5),lty=3,col ="gray")  
legend("bottomleft",c("Curva de Sobrevivencia"),lty=1,col="red")
```

#Gráfico función de riesgo acumulado con ajustes gráficos#

```
plot(.Survfit, conf.int=FALSE, fun="cumhaz",xlab="Tiempo en  
Semestres",ylab="Probabilidad de sobrevida",main="Función de riesgo acumulado  
para Diseño Industrial",lwd=2,col="blue")  
box(lwd=1, col ="black")  
axis(1, seq(0,14,1))  
axis(2, seq(0,1,.1))  
abline( v=c(1,5), col="green")  
legend("bottomright",c("Curva de riesgo acumulado"),lty=1,col="blue")
```

#Desagregado de datos por género#

```
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~ Genero,  
conf.type="log", conf.int=0.95, type="kaplan-meier", error="greenwood",  
data=Datos_DI)  
summary(.Survfit)
```

#Comparación por género#

```
survdif(Surv(Tiempo,censurados) ~ Genero, rho=0, data=Datos_DI)
```

#Gráfica desagregado de datos por género #

```
plot(.Survfit, col=1:2, lty=1:2, mark.time=TRUE)  
legend("bottomleft", legend=c("F","M"), title="Genero", col=1:2, lty=1:2, bty="n")  
title(main="Función de sobrevida por Género", xlab="Tiempo",ylab="Probabilidad  
de sobrevida")
```

#Desagregado de datos por nivel educativo de la madre#

```
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~
Nivel.Educativo.de.la.Madre, conf.type="log", conf.int=0.95, type="kaplan-meier",
error="greenwood",data=Datos_DI)
summary(.Survfit)
```

```
#Comparación por nivel educativo de la madre#
survdif(Surv(Tiempo,censurados) ~ Nivel.educativo.de.la.Madre, rho=0,
data=Datos_DI)
```

```
#Gráfica desagregado de datos por nivel educativo de la madre #
plot(.Survfit, col=1:5, lty=1:5, mark.time=TRUE)
legend("bottomleft", legend=c("BP","BS","TEC","UN"), title="Nivel", col=1:5, lty=1:5,
bty="n")
title(main="Función de sobrevida por Nivel educativo de la madre",
xlab="Tiempo",ylab="Probabilidad de sobrevida")
```

```
#Desagregado de datos por trabaja al presentar el examen de estado#
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~
Trabajaba.al.Presentar.el.Examen.de.Estado, conf.type="log", conf.int=0.95,
type="kaplan-meier", error="greenwood", data=Datos_DI)
summary(.Survfit)
```

```
#Comparación por trabaja al presentar el examen de estado#
survdif(Surv(Tiempo,censurados) ~
Vivienda.Propia,Trabajaba.al.presentar.el.Examen.de.Estado rho=0,
data=Datos_DI)
```

```
#Gráfica desagregado de datos por trabaja al presentar el examen de estado#
plot(.Survfit, col=1:3, lty=1:3, mark.time=TRUE)
legend("bottomleft", legend=c("NO","SI"), title="Trabajaba", col=1:3, lty=1:3,
bty="n")
title(main="Función de sobrevida por Trabajaba al presentar el examen de estado",
xlab="Tiempo",ylab="Probabilidad de sobrevida")
```

```
#Desagregado de datos por vivienda propia#
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~ Vivienda.Propia,
conf.type="log", conf.int=0.95, type="kaplan-meier", error="greenwood",
data=Datos_DI)
summary(.Survfit)
```

```

#Comparación por vivienda propia#
survdif(Surv(Tiempo,censurados) ~ Vivienda.Propia, rho=0, data=Datos_DI)

#Gráfica desagregado de datos por vivienda propia#
plot(.Survfit, col=1:3, lty=1:3, mark.time=TRUE)
legend("bottomleft", legend=c("NO","SI",""), title="Propia", col=1:3, lty=1:3, bty="n")
title(main="Función de sobrevida por Tipo de vivienda",
xlab="Tiempo",ylab="Probabilidad de sobrevida")

#Desagregado de datos por ingreso familiar al presentar examen de estado#
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~
Ingreso.Familiar.al.Presentar.el.Examen.de.Estado, conf.type="log", conf.int=0.95,
type="kaplan-meier", error="greenwood", data=Datos_DI)
summary(.Survfit)

#Comparación por ingreso familiar al presentar examen de estado#
survdif(Surv(Tiempo,censurados) ~
Ingreso.Familiar.al.Presentar.el.Examen.de.Estado, rho=0, data=Datos_DI)

#Gráfica desagregado de datos por ingreso familiar al presentar examen de
estado#
plot(.Survfit, col=1:4, lty=1:4, mark.time=TRUE)
legend("bottomleft", legend=c("AL","BJ","ME"), title="Nivel", col=1:4, lty=1:4,
bty="n")
title(main="Función de sobrevida por Ingresos familiares al presentar el examen de
estado", xlab="Tiempo",ylab="Probabilidad de sobrevida")

###Modelo de cox para deserción###

#Modelo con todas las variables#
CoxModel.DI <- coxph(Surv(Tiempo,censurados) ~ Edad.de.ingreso.al.programa +
Genero + Ingreso.Familiar.al.Presentar.el.Examen.de.Estado +
Nivel.Educativo.de.la.Madre + Numero.de.Hermanos + Puntaje +
Trabajaba.al.Presentar.el.Examen.de.Estado + Vivienda.Propia, +
method="efron", data=Datos_DI)
summary(CoxModel.DI)

##Selección modelo paso a paso##

```

Ae usa la dirección:adelante y el criterio: AIC  
stepwise(CoxModel.DI, direction='forward', criterion='AIC')

#Modelo ajustado#

```
CoxModel.DI.Ajustado <- coxph(Surv(Tiempo,censurados) ~  
Trabajaba.al.Presentar.el.Examen.de.Estado + Edad.de.ingreso.al.programa,  
method="efron", data=Datos_DI)  
summary(CoxModel.DI.Ajustado)
```

#Estimación sobrevida del modelo ajustado#

```
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~ 1, conf.type="log",  
conf.int=0.95, type="kaplan-meier", error="greenwood", data=Datos_DI)  
summary(.Survfit)
```

#Gráfica función de sobrevida del modelo ajustado con ajustes gráficos#

```
plot(.Survfit, conf.int=FALSE,  
mark.time=TRUE,xlab="Semestres",ylab="Probabilidad",main="Función de  
Sobrevida para Diseño Industrial",lwd=2,col="blue")  
box(lwd=1, col ="black")  
axis(1, seq(0,15,1))  
axis(2, seq(0,1,.1))  
abline(v =seq(10,15),lty=2,col ="green")  
legend("bottomright",c("Curva de Sobrevivencia"),lty=1,col="blue")
```

# Gráfica función de riesgo del modelo ajustado con ajustes gráficos #

```
plot(.Survfit, conf.int=FALSE,  
fun="cumhaz",xlab="Semestres",ylab="Probabilidad",main="Función de Riesgo  
Acumulado para Diseño Industrial",lwd=2,col="red")  
box(lwd=1, col ="black")  
axis(1, seq(0,15,1))  
axis(2, seq(0,1,.1))  
abline(v =c(10,15),col ="green")  
legend("bottomright",c("Curva de Riesgo Acumulado"),lty=1,col="red")
```

#Test de riesgos proporcionales#

```
.CoxZPH <- cox.zph(CoxModel.DI.Ajustado)  
.CoxZPH
```

#residuos de cox-snell#

```

estado<-Datos_DI$censurados
mresi<-residuals(CoxModel.DI.Ajustado, type="martingale")
csresi<-estado-mresi
hazard.csresi<-survfit(Surv(csresi,estado)~1,type="fleming-harrington")
plot(hazard.csresi$time , -log(hazard.csresi$surv), xlab='residuos de Cox-Snell',
ylab='riesgo acumulado',lty = 1:4, main="Representación de los residuos de Cox-
Snell")
lines(c(0,5),c(0,5))

```

#####MODELO DE COX PARA GRADUACIÓN#####

#Modelo con todas las variables#

```

CoxModel.DI <- coxph(Surv(Tiempo,censurados) ~ Edad.de.ingreso.al.programa +
Genero + Ingreso.Familiar.al.Presentar.el.Examen.de.Estado +
Nivel.Educativo.de.la.Madre + Numero.de.Hermanos + Puntaje +
Materias.Perdidas+Materias.Repetidas+Trabajaba.al.Presentar.el.Examen.de.Esta
do + Vivienda.Propia, + method="efron", data=Datos_DI)
summary(CoxModel.DI)

```

##Selección modelo paso a paso##

```

Ae usa la dirección:adelante y el criterio: AIC
stepwise(CoxModel.DI, direction='forward', criterion='AIC')

```

#Modelo ajustado#

```

CoxModel.DI.Ajustado <- coxph(Surv(Tiempo,censurados) ~ Materias.Perdidas,
method="efron", data=Datos_DI)
summary(CoxModel.DI.Ajustado)

```

#Estimación sobrevida del modelo ajustado#

```

.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~ 1, conf.type="log",
conf.int=0.95, type="kaplan-meier", error="greenwood", data=Datos_DI)
summary(.Survfit)

```

#Gráfica función de sobrevida del modelo ajustado con ajustes gráficos#

```

plot(.Survfit, conf.int=FALSE,
mark.time=TRUE,xlab="Semestres",ylab="Probabilidad",main="Función de
Sobrevida para Diseño Industrial",lwd=2,col="blue")
box(lwd=1, col ="black")
axis(1, seq(0,15,1))

```

```

axis(2, seq(0,1,.1))
abline(v =seq(10,15),lty=2,col ="green")
legend("bottomright",c("Curva de Supervivencia"),lty=1,col="blue")

# Gráfica función de riesgo del modelo ajustado con ajustes gráficos #
plot(.Survfit, conf.int=FALSE,
fun="cumhaz",xlab="Semestres",ylab="Probabilidad",main="Función de Riesgo
Acumulado para Diseño Industrial",lwd=2,col="red")
box(lwd=1, col ="black")
axis(1, seq(0,15,1))
axis(2, seq(0,1,.1))
abline(v =c(10,15),col ="green")
legend("bottomright",c("Curva de Riesgo Acumulado"),lty=1,col="red")

#Test de riesgos proporcionales#
.CoxZPH <- cox.zph(CoxModel.DI.Ajustado)
.CoxZPH

#residuos de cox-snell#
estado<-Datos_DI$censurados
mresi<-residuals(CoxModel.DI.Ajustado, type="martingale")
csresi<-estado-mresi
hazard.csresi<-survfit(Surv(csresi,estado)~1,type="fleming-harrington")
plot(hazard.csresi$time , -log(hazard.csresi$surv), xlab='residuos de Cox-Snell',
ylab='riesgo acumulado',lty = 1:4, main="Representación de los residuos de Cox-
Snell")
lines(c(0,5),c(0,5))

#####
#####INGENIERÍA ELECTROMECÁNICA#####
#####

## Pasamos la base de datos de xls de Ingeniería Electromecánica al R-
Commander##
Datos_ING <- sqlQuery(channel = 1, select * from [ING$])
names(Datos_ING) <- make.names(names(Datos_ING))

#####DESCRIPCIÓN DE LOS DATOS INGENIERÍA ELECTROMECÁNICA#####

```

```
####Resumen Univariado####  
summary(Datos_ING)
```

```
####Resumen Bivariado####  
##Variables categóricas##
```

```
#Variables: Género y Desertar#  
.Table <- xtabs(~Genero+Deserto, data=Datos_ING)  
Género <- xtabs(~Genero+Deserto, data=Datos_ING)  
barplot(Género,scale="percentages",main="Estado del Estudiante Vs  
Género",xlab="Estado del Estudiante",ylab="Porcentaje de  
Estudiantes",names.arg=c("Desertor","No Desertor"),  
beside=TRUE,col=topo.colors(2))  
legend(c("Femenino","Masculino"),fill=topo.colors(2))
```

```
#Variables: Vivienda propia y Desertar#  
.Table <- xtabs(~Vivienda.Propia+Deserto, data=Datos_ING)  
Vivienda <- xtabs(~Vivienda.Propia+Deserto, data=Datos_ING)  
barplot(Vivienda,scale="percentages",main="Estado del Estudiante Vs Vivienda  
Propia",xlab="Estado de vivienda del Estudiante al presentar el examen de  
Estado",ylab="Porcentaje de Estudiantes",names.arg=c("Desertor","No Desertor"),  
beside=TRUE,col=heat.colors(3))  
legend("topleft",c("Posee","Carece","Sin Clasificar"),fill=heat.colors(3))
```

```
#Variables: Trabajaba al presentar examen de estado y Desertar#  
.Table <- xtabs(~Trabajaba.al.Presentar.el.Examen.de.Estado+Deserto,  
data=Datos_ING)  
Trabajo <- xtabs(~Trabajaba.al.Presentar.el.Examen.de.Estado+Deserto,  
data=Datos_ING)  
barplot(Trabajo,scale="percentages",main="Estado del Estudiante Vs  
Trabajaba",xlab="Estado de trabajo del Estudiante al presentar el examen de  
Estado",ylab="Porcentaje de Estudiantes",names.arg=c("Desertor","No  
Desertor"),beside=TRUE,col=topo.colors(3))  
legend("topright",c("No","Sin Clasificar"),fill=topo.colors(2))
```

```
#Variables: Nivel educativo de la madre al presentar examen de estado y  
Desertar#  
.Table <- xtabs(~Nivel.Educativo.de.la.Madre+Deserto, data=Datos_ING)  
Nivel.educa <- xtabs(~Nivel.Educativo.de.la.Madre+Deserto, data=Datos_ING)
```

```

barplot(Nivel.educa,scale="percentages",main="Estado del Estudiante Vs Nivel
Educativo de la Madre",xlab="Nivel educativo de la madre al presentar el examen
de Estado",ylab="Porcentaje de Estudiantes",names.arg=c("Desertor","No
Desertor"),beside=TRUE,col=topo.colors(5))
legend(4,38,c("Básica Primaria","Básica Secundaria","Sin
Clasificar","Tecnológico","Universitario"),cex=0.95,fill=topo.colors(5))

```

```

#Variables: Ingreso familiar al presentar examen de estado y Desertar#
.Table <- xtabs(~Ingreso.Familiar.al.Presentar.el.Examen.de.Estado+Deserto,
data=Datos_ING)
Ingreso <- xtabs(~Ingreso.Familiar.al.Presentar.el.Examen.de.Estado+Deserto,
data=Datos_ING)
barplot(Ingreso,scale="percentages",main="Estado del Estudiante Vs Ingreso
Familiar",xlab="Ingreso Familiar al presentar el examen de
Estado",ylab="Porcentaje de Estudiantes",names.arg=c("Desertor","No
Desertor"),beside=TRUE,col=rainbow(6))
legend("topright",c("[0-1)","[1-2)","[10-)","[2-3)","[3-5)","[5-7)","Sin
Clasificar"),fill=rainbow(7))

```

##Correlación de Pearson de variables continuas en Ingeniería Electromécanica##

```

#Variables: Tiempo hasta desertar, Edad de Presentación del Examen de Estado#
with(Datos_ING,cor.test(Tiempo.hasta.que.deserto,Edad.de.Presentación.del.Exa
men.de.Estado,alternative="two.sided",method="pearson",exact=NULL,conf.level=
0.95))

```

```

#Variables: Tiempo hasta desertar, Número de hermanos#
with(Datos_ING,cor.test(Tiempo.hasta.que.deserto,Numero.de.Hermanos,alternati
ve="two.sided",method="pearson",exact=NULL,conf.level=0.95))

```

```

#Variables: Tiempo hasta desertar, Puntaje#
with(Datos_ING,cor.test(Tiempo.hasta.que.deserto,Puntaje,alternative="two.sided"
,method="pearson",exact=NULL,conf.level=0.95))

```

```

#Variables: Tiempo hasta desertar, Materias perdidas#
with(Datos_ING,cor.test(Tiempo.hasta.que.deserto,Materias.Perdidas,alternative="
two.sided",method="pearson",exact=NULL,conf.level=0.95))

```

```

#Variables: Tiempo hasta desertar, Materias repetidas#

```



```
with(Datos_ING,cor.test(Tiempo.hasta.que.deserto,Materias.Repetidas,alternative
="two.sided",method="pearson",exact=NULL,conf.level=0.95))
```

```
#Desagregado de datos por nivel educativo de la madre#
```

```
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~
Nivel.Educativo.de.la.Madre, conf.type="log", conf.int=0.95, type="kaplan-meier",
error="greenwood",data=Datos_ING)
summary(.Survfit)
```

```
#Comparación por nivel educativo de la madre#
```

```
survdif(Surv(Tiempo,censurados) ~ Nivel.educativo.de.la.Madre, rho=0,
data=Datos_ING)
```

```
#Gráfica desagregado de datos por nivel educativo de la madre#
```

```
plot(.Survfit, col=1:5, lty=1:5, mark.time=TRUE)
legend("bottomleft", legend=c("BP","BS","","TEC","UN"), title="Nivel", col=1:5,
lty=1:5, bty="n")
title(main="Función de sobrevivencia por Nivel educativo de la madre",
xlab="Tiempo",ylab="Probabilidad de sobrevivencia")
```

```
#Desagregado de datos por vivienda propia#
```

```
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~ Vivienda.Propia,
conf.type="log", conf.int=0.95, type="kaplan-meier", error="greenwood",
data=Datos_ING)
summary(.Survfit)
```

```
#Comparación por vivienda propia#
```

```
survdif(Surv(Tiempo,censurados) ~ Vivienda.Propia, rho=0, data=Datos_ING)
```

```
#Gráfica desagregado de datos por vivienda propia#
```

```
plot(.Survfit, col=1:3, lty=1:3, mark.time=TRUE)
legend("bottomleft", legend=c("NO","SI",""), title="Propia", col=1:3, lty=1:3, bty="n")
title(main="Función de sobrevivencia por Tipo de vivienda",
xlab="Tiempo",ylab="Probabilidad de sobrevivencia")
```

```
#Desagregado de datos por ingreso familiar al presentar examen de estado#
```

```
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~
Ingreso.Familiar.al.Presentar.el.Examen.de.Estado, conf.type="log", conf.int=0.95,
type="kaplan-meier", error="greenwood", data=Datos_ING)
```

```
summary(.Survfit)
```

```
#Comparación por ingreso familiar al presentar examen de estado#  
survdif(Surv(Tiempo,censurados) ~  
Ingreso.Familiar.al.Presentar.el.Examen.de.Estado, rho=0, data=Datos_ING)
```

```
#Gráfica desagregado de datos por ingreso familiar al presentar examen de  
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~  
Ingreso.Familiar.al.Presentar.el.Examen.de.Estado, conf.type="log", conf.int=0.95,  
type="kaplan-meier", error="greenwood", data=Datos_ING)  
summary(.Survfit)
```

```
#Comparación por ingreso familiar al presentar examen de estado#  
survdif(Surv(Tiempo,censurados) ~  
Ingreso.Familiar.al.Presentar.el.Examen.de.Estado, rho=0, data=Datos_ING)
```

```
#Gráfica desagregado de datos por ingreso familiar al presentar examen de  
estado#  
plot(.Survfit, col=1:4, lty=1:4, mark.time=TRUE)  
legend("bottomleft", legend=c("AL","BJ","ME"), title="Nivel", col=1:4, lty=1:4,  
bty="n")  
title(main="Función de sobrevivencia por Ingresos familiares al presentar el examen de  
estado", xlab="Tiempo", ylab="Probabilidad de sobrevivencia")
```

```
#####MODELO DE COX PARA DESERCIÓN#####
```

```
#Diagrama de LEXIS#  
LL <- Lexis.diagram( age=c(-1,15), date=c(2003,2015), entry.age= Edad.ingreso,  
exit.age= Edad.salida, birth.date=Cohorte, fail=(censurados), lwd.life=1,  
cex.fail=0.8, col.fail=c("blue","red"), alab="Tiempo de deserción", dlab="Cohorte",  
int=c(1,1),data=Datos_ING)  
abline( v=c(2004,2015), col="blue")  
abline( h=1, col="green")  
abline( h=5, col="green")
```

```
#Identifica personas según entrada y salida#  
text( LL$exit.date, LL$exit.age, paste(1:nrow(LL)), col="Black", font=10, adj=c(0,1))
```

```
#Modelo con todas las variables#
CoxModel.ING <- coxph(Surv(Tiempo,censurados) ~ Edad.de.ingreso.al.programa
+ Genero + Ingreso.Familiar.al.Presentar.el.Examen.de.Estado +
Nivel.Educativo.de.la.Madre + Numero.de.Hermanos + Puntaje +
Trabajaba.al.Presentar.el.Examen.de.Estado + Vivienda.Propia, +
method="efron", data=Datos_ING)
summary(CoxModel.ING)
```

```
##Selección modelo paso a paso##
Ae usa la dirección:adelante y el criterio: AIC
stepwise(CoxModel.ING, direction='forward', criterion='AIC')
```

#####MODELO DE COX PARA GRADUACIÓN#####

```
#Modelo con todas las variables#
CoxModel.ING <- coxph(Surv(Tiempo,censurados) ~ Edad.de.ingreso.al.programa
+ Genero + Ingreso.Familiar.al.Presentar.el.Examen.de.Estado +
Nivel.Educativo.de.la.Madre + Numero.de.Hermanos + Puntaje +
Materias.Perdidas+Materias.Repetidas+Trabajaba.al.Presentar.el.Examen.de.Esta
do + Vivienda.Propia, + method="efron", data=Datos_ING)
summary(CoxModel.ING)
```

```
##Selección modelo paso a paso##
Ae usa la dirección:adelante y el criterio: AIC
stepwise(CoxModel.ING, direction='forward', criterion='AIC')
```

```
#Modelo ajustado#
CoxModel.ING.Ajustado <- coxph(Surv(Tiempo,censurados) ~ Materias.Perdidas,
method="efron", data=Datos_ING)
summary(CoxModel.ING.Ajustado)
```

```
#Estimación sobrevida del modelo ajustado#
.Survfit <- survfit(Surv(Tiempo, censurados, type="right") ~ 1, conf.type="log",
conf.int=0.95, type="kaplan-meier", error="greenwood", data=Datos_ING)
summary(.Survfit)
```

```
#Gráfica función de sobrevida del modelo ajustado con ajustes gráficos#
```

```

plot(.Survfit, conf.int=FALSE,
mark.time=TRUE,xlab="Semestres",ylab="Probabilidad",main="Función de
Sobrevivencia para Ingeniería Electromécanica",lwd=2,col="blue")
box(lwd=1, col ="black")
axis(1, seq(0,16,1))
axis(2, seq(0,1,.1))
abline(v =seq(10,16),lty=2,col ="green")
legend("bottomright",c("Curva de Sobrevivencia"),lty=1,col="blue")

# Gráfica función de riesgo del modelo ajustado con ajustes gráficos #
plot(.Survfit, conf.int=FALSE,
fun="cumhaz",xlab="Semestres",ylab="Probabilidad",main="Función de Riesgo
Acumulado para Ingeniería Electromécanica",lwd=2,col="red")
box(lwd=1, col ="black")
axis(1, seq(0,16,1))
axis(2, seq(0,1,.1))
abline(v =c(10,16),col ="green")
legend("bottomright",c("Curva de Riesgo Acumulado"),lty=1,col="red")

#Test de riesgos proporcionales#
.CoxZPH <- cox.zph(CoxModel.ING.Ajustado)
.CoxZPH

#residuos de cox-snell#
estado<-Datos_ING$censurados
mresi<-residuals(CoxModel.ING.Ajustado, type="martingale")
csresi<-estado-mresi
hazard.csresi<-survfit(Surv(csresi,estado)~1,type="fleming-harrington")
plot(hazard.csresi$time , -log(hazard.csresi$surv), xlab='residuos de Cox-Snell',
ylab='riesgo acumulado',lty = 1:4, main="Representación de los residuos de Cox-
Snell")
lines(c(0,5),c(0,5))

```

## ANEXO D. FORMATO DE REGISTRO DE DATOS

Teniendo en cuenta que el análisis efectuado en este proyecto indicó un pobre ajuste, es importante indicar algunas otras variables que podrían estar afectando el riesgo de deserción y la graduación. Lo anterior con el fin de que cada programa recoja la información pertinente a sus estudiantes y a futuro la utilice para construir modelos de sobrevivencia con una mayor bondad de ajuste.

Las variables que se proponen se dividen en 4 grupos, así:

### I. Información Personal

- Semestre de ingreso: año del semestre en el que un estudiante se matriculó al programa, ejemplo: 2016-I, significa primer semestre del 2016.
- Código: Valor numérico asignado por registro en el momento de la matrícula.
- *Género*, los valores que toma son: Masculino (M), Femenino (F).
- *Edad*: valor numérico de la edad en la que se matriculó en el programa.
- Estado Civil: Soltero(a), Casado(a), Divorciado(a), Viudo(a)
- Si tiene hijos, indicar el número de estos.
- Si tiene hermanos, señalar el número de hermanos
- Posición entre hermanos: No tiene (0), primer lugar (1), segundo (2), así sucesivamente.
- Estado de la salud Física: Bueno, Malo, Regular.
- *Población Vulnerable*: Desplazado (DESPL), Afrodescendiente (AFRO), LGTBI, Ninguno.
- *Embarazos no planeados*: si ha experimentado tales, Si o No.
- *Calamidad y/o problema doméstico* los valores que toma: Si presentó algunos de los siguientes eventos, Muerte de los padres (MP), Muerte de un Hijo (MH), Ninguno.
- *Expectativas no satisfechas*: Percepciones de estudiante antes de ingresar al programa, posibles valores que puede tomar: no le gusta el programa(NGPR), no era como se lo esperaba(NESP), es muy duro el cambio del colegio a la Universidad(CAMB\_U)

- *Discapacidad*: si el estudiante manifiesta tenerla, posibles valores a tomar: Física, Sensorial: dentro de esta se encuentran (auditiva, visual), psíquica, intelectual o mental.

## **II. Información Socioeconómica**

- Situación Laboral del estudiante al momento de ingresar a la universidad: valores que toma la variable: Si, No.
- ¿Cuánto es el nivel de ingresos?: Bajo (0, 1 SMMLV], Medio (1,3 SMMLV], Alto (3, en adelante], SMMLV: Salario Mínimo Mensual Legal Vigente.
- ¿Con quién vive? Madre y padre (1), Madre o padre (2), Otro familiar (3), ningún familiar (4)
- Si se tienen personas a cargo, indicar el número de estas
- Nivel del Sisben, indica el nivel de este, valores que toma (1,2,3)
- Clasificación del Estrato socioeconómico los valores que toma son (1,2,3,4,5)
- Nivel de Ingresos de los Padres, indica cual es el ingreso familiar de los padres, y los valores que puede tener: Bajo (0, 1 SMMLV], Medio (1,3 SMMLV], Alto (3, en adelante]
- Máximo Nivel de estudios alcanzado por el Padre, estos valores son: Ninguno, Primaria, Básica, Media, Técnico, Universitario-Superior.
- Máximo Nivel de estudios alcanzado por el Madre, estos valores son: Ninguno, Primaria, Básica, Media, Técnico, Universitario-Superior.
- Ocupación del padre: Jubilado, hogar, estudiante, busca empleo (Baja), trabajador independiente, empleado (Media), empresario, administrador, gerente, profesional independiente (Alta).
- Ocupación de la madre: Jubilada, hogar, estudiante, busca empleo (Baja), trabajador independiente, empleada (Media), empresaria, administradora, gerente, profesional independiente (Alta).
- Tipo de vivienda, Si es propia o No.
- Número de Hermanos, indica la cantidad de hermanos que tiene.

- Tipo de relación con los padres, la opinión del estudiante y su relación con sus progenitores, puede tomar los siguientes valores: Buena, mala, regular.


### **III. Variables Académicas**

- *Tiempo de ingreso a la Universidad:* indica el número de semestres que el estudiante demoró en ingresar a la Universidad.
- Puntaje estandarizado de la prueba saber 11.
- Tipo de Colegio: indica el tipo de colegio de egreso del estudiante, tiene dos categorías Público o Privado.
- *Rendimiento académico durante el Colegio:* Cómo considera que fue este, puede ser: Excelente, Sobresaliente, Aceptable, Insuficiente.
- *¿Ha recibido algún tipo de orientación vocacional?* Indique Si o No.
- Énfasis del colegio de egreso de secundaria del estudiante: técnico agropecuario, comercial=1, académico=2, turístico=3, técnico=4.
- Orientación Vocacional previa recibida antes de ingresar a la universidad. Tiene dos categorías posibles: Sin O.V.=1 si no recibió ninguna orientación vocacional, O.V.=1 si recibió orientación.
- Empezar Universidad: Esta variable cuenta con cuatro categorías. E1: es la primera vez que el alumno comienza un estudio superior, COC: si además de la carrera iniciada, continua otra carrera, TOC: en caso en que el estudiante, al iniciar esta carrera, ya haya terminado previamente algún otro estudio superior. Y por último, AC: en caso que abandono una carrera anteriormente.

### **IV. Información Institucional**

- ¿Ha recibido beca o financiamiento de la Universidad? Si, No.
- ¿El programa académico que cursa, cuenta con registro calificado? Si, No.
- ¿Considera que usted tiene una buena relación con los docentes? Si, No.
- ¿Considera que usted tiene una buena relación con sus compañeros? Si, No.
- ¿Ha iniciado otro programa académico en esta Institución u otra?, Indicar Si o no, además. Mencione ¿cuál?\_\_\_\_\_

## VISTA DE LOS ENCABEZADOS DE LA HOJA DE EXCEL

														
INFORMACIÓN PERSONAL														
N°	SEMESTRE DE INGRESO	CÓDIGO	GÉNERO	EDAD	ESTADO CIVIL	Nº DE HERMANOS	POSICIÓN ENTRE HERMANOS	Nº HIJOS	ESTADO SALUD FISICA	POBLACIÓN VULNERABLE	EMBARAZOS NO PLANEADOS	CALAMIDAD Y/O PROBLEMAS DOMÉSTICOS	EXPECTATIVAS NO SASTISFECHAS	DISCAPACIDAD
1														

### FORMATO DE CAPTURA DE DATOS - VARIALES NO ESTATICAS

INFORMACIÓN SOCIECONOMICA														
SITUACIÓN LABORAL	CUENTA PROPIA	NIVEL DE INGRESOS ESTUDIANTE	PERSONAS CON QUIEN VIVE	# PERSONAS A CARGO	NIVEL DEL SISBÉN	ESTRATO	NIVEL DE INGRESOS PADRES	NIVEL EDUCATIVO PADRE	NIVEL EDUCATIVO MADRE	OCUPACIÓN DEL PADRE	OCUPACIÓN DE LA MADRE	TIPO DE VIVIENDA	NUMERO DE HERMANOS	TIPO DE RELACIÓN CON PADRES



ACADÉMICAS										INSTITUCIONAL					
TIEMPO DE INGRESO A LA UNIV.	PUNTAJE P.SABER	TIPO COLEGIO	TIPO PROGRAMA ACADÉMICO	RENDIMIENTO ACADEMICO EN EL COLEGIO	ORIENTACION VOCACIONAL	Énfasis colegio de egreso	Orientación vocacional previa	Empezar Universidad	INSATISFACCIÓN CON EL PROGRAMA	NORMATIVAS ACADÉMICAS	BECAS/FINANCIAMIENTO	CALIDAD DEL PROGRAMA	RELACIÓN CON PROFESORES	RELACIÓN CON ESTUDIANTES	PROVIENE DE OTRO PROGRAMA